DOCUMENT RESUME

ED 263 165 TM 850 650

AUTHOR Kahan, James P.; And Others
TITLE Individual Characteristics and Unit Performance. A Review of Research and Methods.
INSTITUTION Office of the Assistant Secretary of Defense for Manpower, Installations and Logistics (DOD), Washington, DC.
REPORT NO ISBN-0-8330-0604-5; R-3194-MIL
PUB DATE Feb 85
CONTRACT MDA903-83-C-0047
NOTE 123p.
PUB TYPE Information Analyses (070)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Adults; *Armed Forces; *Group Dynamics; Group Structure; *Individual Characteristics; Interpersonal Communication; *Job Performance; Literature Reviews; Peer Influence; *Performance Factors
IDENTIFIERS Combat

ABSTRACT

This report reviews existing studies of the determinants of group performance in an attempt to understand how personnel characteristics of units affect the effectiveness and efficiency with which those units perform their missions. As defined for this project, unit performance is the aggregate behavior of personnel in a unit. This review provides a broad survey of the results that have been found in both the civilian (specifically, team sports) and the military literature that might have an application to small units engaged in combat arms. It discusses these results in terms of the following categories of predictors of group performance: (1) characteristics of individuals; (2) leadership; (3) group structure; (4) group processes; and (5) training techniques. The reliability and validity of unit performance measurements are described in the appendix. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

ED263165

Individual Characteristics and Unit Performance

A Review of Research and Methods

James P. Kahan, Noreen Webb,
Richard J. Shavelson, Ross M. Stolzenberg

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Cockrell, T.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM 830 650

Rand

The research described in this report was sponsored by the Office of the Assistant Secretary of Defense/Manpower, Installations and Logistics under Contract MDA903-83-C-0047.

Library of Congress Cataloging in Publication Data

Main entry under title:

Individual characteristics and unit performance.

"Prepared for the Office of the Assistant Secretary of Defense/Manpower, Installations, and Logistics."
"R-3194-MIL."
"February 1985."
Bibliography: p.
1. Performance standards—Research. 2. Work groups—Research. I. Kahan, James P. II. Rand Corporation. III. United States. Office of the Assistant Secretary of Defense (Manpower, Installations, and Logistics)
HF5549.5.P35I53 1985 658.4'03 84-18216
ISBN 0-8330-0604-5

The Rand Publication Series: The Report is the principal publication documenting and transmitting Rand's major research findings and final research results. The Rand Note reports other outputs of sponsored research for general distribution. Publications of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

Published by The Rand Corporation

BEST COPY AVAILABLE

R-3194-MIL

Individual Characteristics and Unit Performance

A Review of Research and Methods

James P. Kahan, Noreen Webb,
Richard J. Shavelson, Ross M. Stolzenberg

February 1985

Prepared for the
Office of the Assistant Secretary of Defense/
Manpower, Installations and Logistics

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

PREFACE

This report presents the results of a review of the psychological literature to determine the characteristics of individuals and groups that predict the quality of performance of small groups on tasks requiring ability and skill. The research, which was conducted in Rand's Defense Manpower Research Center, was sponsored by the Office of the Assistant Secretary of Defense for Manpower, Installations, and Logistics under Contract No. MDA903-83-C-0047.

These findings and their implications for policy and future research are intended for a diverse audience, including government policymakers and the social scientific research community.

SUMMARY

In this review, we examined the nature of unit performance and searched for predictors of quality performance. The search encompassed the topics of the characteristics of individuals, characteristics of groups, leadership characteristics, group structure, group processes, and team training techniques. Because unit performance is so broadly defined, much of the research yielded ambiguous or seemingly contradictory prescriptions; put another way, there are many variables that interact to determine unit performance, so that without a good specification of these variables, consistent prediction is unlikely. Within this context of complexity, though, there did emerge some consistent patterns.

There is general agreement that objective measures of performance, keyed to small behavioral segments performed by working groups, will yield more reliable and valid results than subjective, global measures of performance. Moreover, feedback in terms of such measures produces more improvement in performance than more general feedback. Therefore, efforts to specify task performance in small behavioral units, which is an ongoing effort in the development of Army training techniques, should be continued and widened in scope.

A major distinction between unit environments is whether they are interactive or coactive. Interactive environments call for individual duties that are collaborative and involve joint action, whereas coactive environments are those in which group productivity is a function of separate, albeit coordinated, individual effort. Most unit performance tasks in the Army are more interactive than they are coactive. The distinction between the two types is important because predictors of unit performance are more often than not dependent on whether the task is interactive or coactive.

A number of studies using general individual ability, individual task proficiency, and the heterogeneity of group proficiency as predictors have shown a common pattern of predictiveness on unit performance. For coactive tasks, the higher the ability of individual group members, or the greater the heterogeneity of the group, the better was performance, particularly in the learning stages of any task. Over a number of studies of coactive tasks, from one-quarter to one-half of the variation in performance quality could be attributable to the ability of the members. The more routine the task, the less greater practice affected ability. On the other hand, with interactive tasks, the effect of ability was reduced, if present at all, and outcomes were much more task-

specific. For some interactive tasks, there is a "bottleneck" effect, where performance is more determined by the least-able member, while for other tasks, there is an opposite effect, where the most-able member predominates and determines performance. Which of these effects will obtain depends on the specific nature of the task. For tasks in which members may easily replace each others' roles, the more-able members can perform multiple functions, and their ability will determine performance. For tasks in which there is little role flexibility, the least-able member determines performance.

It is almost tautologically true that the higher a person's motivation, the better will be his performance. However, this generality must be qualified by the research evidence that what motivates individuals to perform in any given task is not obvious and may even be counterintuitive. For any particular program, a brief investigation to ascertain the specific motivations of the unit members, perhaps in the form of focus groups or interviews, should precede the establishment of a reward structure.

There are a great number of studies examining the effect of the personalities of group members and group leaders on group productivity. However, these studies have not followed any systematic pattern of investigation, and together do not offer any recommendations for assembling units so as to improve performance. However, a systematic research program on leader behavior has identified a number of behaviors that lead to more effective leadership and better unit performance. Among those behaviors are an emphasis on performance, maintenance of well-defined roles for group members, attentive management control, and an adviser or counsellor for supervised personnel. Leadership training programs that teach those skills should improve unit performance.

The homogeneity of the unit may be a factor influencing performance. Homogeneity of ability may either help or hinder interactive tasks, as was discussed above. Homogeneity in terms of socioeconomic, demographic, or personality characteristics presents a somewhat perplexing picture. On the one hand, a number of studies have shown that homogeneity of such characteristics prevents the formation of disruptive cliques and leads to better performance. On the other hand, a different set of studies has shown that groups that are very cohesive in the sense of liking each other very much attend more to the socioemotional aspects of the group relationship to the detriment of performance, and so perform at lower levels. The two findings are contradictory because groups that are homogeneous tend to have more liking among group members.

There is clear evidence that a common orientation toward task productivity is associated with superior performance. However, the causal direction of that association is not firmly established. The weight of present evidence tilts more toward the hypothesis that successful unit experience engenders feelings of cohesiveness rather than cohesiveness producing successful experience. Moreover, too much affective cohesion, or group emotional solidarity, might interfere with the critical appraisal of performance that is needed to maintain quality. Therefore, at present, there is little incentive for programmatic measures to improve group cohesiveness.

Finally, our review of the work on team training techniques supports current efforts by armed forces investigators. Feedback on performance, both on individual and group levels and in the form of information about specific behavioral segments, improves performance. Simulation exercises, especially those employing new high-technology devices, provide surrogate battlefield experience that aids performance. There is also a need for training in communication, so that team members can communicate efficiently and effectively. The motivation of team members can be affected by appropriate training and inducements; more research on effective techniques is needed. Finally, the complex task of team members, in which balances must be struck in terms of specialized roles vs. procedural flexibility, individual initiative vs. team coordination, and "rational" task orientation vs. esprit de corps, requires further research within military settings; analyses of extant preliminary studies indicate that this line of research promises to yield practical results.

ACKNOWLEDGMENTS

The authors gratefully acknowledge their indebtedness to several people who have contributed to this effort: Bruce Orvis, Roberta Smith, Cathleen Stasz, and Gail Zellman, all of The Rand Corporation; and Leigh Burstein of the University of California at Los Angeles.

CONTENTS

PREFACE iii

SUMMARY v

ACKNOWLEDGMENTS ix

FIGURE AND TABLES xiii

ACRONYMS xv

Section

I. INTRODUCTION 1
- Background 1
- Overview 3

II. DEFINITIONS 7
- Units of Analysis 7
- Defining a Unit 7
- Defining Performance 9
- Defining Group Tasks 12

III. PREDICTING GROUP PERFORMANCE FROM CHARACTERISTICS OF INDIVIDUALS 15
- General Ability 15
- Individual Task Proficiency Prior to Teamwork 19
- Personality and Motivation 23
- Summary 27

IV. LEADERSHIP 29
- Ability of the Leader 30
- Personality of the Leader 33
- Leader Behavior and Leadership Styles 35
- Summary 39

V. GROUP STRUCTURE 40
- Size and Turbulence 40
- Distribution of General Ability 43
- Task-Related Abilities and Proficiency in the Task 46
- Homogeneity of Personality and Cognitive Styles 50
- Interpersonal Compatibility 55
- Summary 63

VI. GROUP PROCESSES 64
Group Cohesiveness 65
Interpersonal Attraction 67
Summary 69

VII. TRAINING TECHNIQUES 70
Feedback About Performance 70
Training Through Simulation 72
Motivation in Training 73
Team-Oriented Training 73
Summary 77

VIII. CONCLUSIONS 79

Appendix: THE RELIABILITY AND VALIDITY OF UNIT PERFORMANCE MEASUREMENTS 83

REFERENCES 97

FIGURE

1. Fiedler's Contingency Model 38

TABLES

A.1. Estimates of Variance Components for a One-Facet, $K \times O$ Design 86
A.2. Expected Mean Squares in a $C \times P(C) \times Cr(PC) \times O$ Generalizability Design 91
A.3. Variance Components from the Table VIII Dataset 91

ACRONYMS

ARTEP	ARmy Training and Evaluation Program
ASVAB	Armed Services Vocational Aptitude Battery
CO	COmbat potential—a scale of the ASVAB
CONUS	CONtinental United States
COTEAM	Combat Operations Training Effectiveness Analysis Model
FIRO-B	Fundamental Interpersonal Relationship Orientation-Behavior—a personality test
FORSCOM	FORceS COMmand of the U.S. Army
G	Generalizability theory—an analytic technique for examining the reliability of data
GCT	General Classification Test—a measure of ability
LPC	Least Preferred Coworker score—a measure of leadership style
MILES	Multiple Integrated Laser Engagement System—a modern, high-technology training simulation exercise
MOS	Military Occupational Specialty
ORTT	Operational Readiness Training Test—a field test
REALTRAIN	A REAListic TRAINing simulation exercise
SAM	Surface-to-Air Missile
SAT	Scholastic Aptitude Test
SQT	Skill Qualification Test—an Army performance measure
TAT	Thematic Apperception Test—a personality test
TRADOC	TRAining and DOCtrine command of the U.S. Army

I. INTRODUCTION

BACKGROUND

This study is an initial effort to understand how characteristics of individuals influence the effectiveness and efficiency with which the military units to which they belong perform their missions. This project was originally motivated by Congressional interest in the relationship between enlistment standards and military performance. Congress, like the Services and the Office of the Secretary of Defense, observed the continuing difficulty of recruiting high-quality enlisted personnel, and wished to know the extent to which different ability mixes in the enlisted force would produce differences in the capabilities of the Armed Services to perform their missions. There is much research on the relationship between attributes of individuals and their performance of one-person tasks. But modern military combat is normally a group task, and at the time this study was undertaken, there appeared to be very little research on the way that characteristics of individual members of a group affect the performance of tasks by the group as a whole. Therefore, this research was undertaken with three goals in mind:

- Systematic review of knowledge about the relationship between individual attributes of group members and the efficiency and effectiveness with which their group performs collective tasks.
- Identification and evaluation of potential sources of data on the relationship between group performance and the attributes of individual group members.
- Acquisition of performance data analyses of the reliability and validity of performance measures, and statistical modelling of the relationship between group performance and the characteristics of individual group members.

The third of these goals was frustrated by concern over the confidentiality of performance data. The second goal became, over time, unimportant to the client for whom this research was undertaken. And so the first goal, assessment of current knowledge of the relationship between individual characteristics and group performance, became the single focus of this project.

Before proceeding, we note in more detail that the relationship between personnel characteristics and unit performance has broad relevance to a variety of military personnel management issues. This relevance is illustrated by a few examples:

- Obtaining high-aptitude accessions has remained a problem for the all-volunteer force. If all military tasks were individual activities, there would be a relatively simple, monotonic relationship between the aptitude of individuals and the performance of tasks. But many activities, including combat, are group tasks. So the aptitude mix of a unit, rather than the simple aptitude of an individual, becomes relevant to considerations of aptitude and performance. Answers to certain key questions about unit performance would shed light on the relationship between aptitude mix and performance, and would indicate the feasibility of manipulating the ability mix of a unit's members to enhance the group's performance, even with a fixed distribution of abilities in the force. These questions include: Does a single high-aptitude member of a unit make up for a low average level of aptitude in a unit? During unit training, does the presence of a single high-aptitude member of the unit affect the learning curve of the unit as a whole? Does high or low variance in the individual aptitudes of unit members affect the performance of the unit as a whole? Does a single low-aptitude member of a unit drag down performance of the entire unit?

Similarly, the retention of experienced personnel is believed to be important for mission- as well as cost-effectiveness. Just as the ability mix of a group is important to consider, the experience mix of a group may be an important variable in maximizing group performance.

- Keeping individuals together in working units is logistically complex, expensive, and reduces management flexibility. Yet there is reason to believe that keeping units together improves their task performance, and may even affect the propensities of their members to terminate military service. Balancing benefits and costs of keeping units together requires estimates of the relationship between a unit's length of time together and its performance as a unit. For example: To what extent does the length of a unit's experience together affect the unit's performance? To what extent, if any, do other factors such as training, aptitude, or task complexity affect the relationship between time together and performance? Does longer experience as a working unit compensate for lower experience levels of individ-

uals in the unit? Does time together as a unit have less impact on unit performance when individual unit members have high task skill levels (e.g., Skill Qualification Test (SQT) scores) than when individual members have moderate or low task skill levels?

- In theory, leadership can make up for individual deficiencies in unit members' ability or experience. In practice, it would be useful to gain some quantitative measure of the ways that specific characteristics of unit leaders affect the performance of their units. For example: What are the effects on a unit's performance of the unit leader's length of experience with that unit, length of total military experience, length of experience as a commander of other units, and mental ability? Do high levels of experience, mental ability, or skill among unit members make up for low levels of unit commander experience? Do high levels of unit commander ability and experience make up for low levels of ability or experience among his subordinates? To what extent do these effects vary with the type of task performed by the unit?

These are just a few examples of the policy questions which can be addressed by information about the relationship between characteristics of individuals and the performance of the combat units to which they belong. We believe that these questions illustrate the importance of knowing what determines unit performance, and of applying that information to the complex problems of accession policy, training, and force management.

The present study therefore reviews existing studies of the determinants of group performance, in an attempt to understand how personnel characteristics of units affect the effectiveness and efficiency with which those units perform their missions. As defined for this project, unit performance is the aggregate behavior of personnel in a unit. This definition excludes nonpersonnel characteristics such as equipment, weapons, or other logistics associated with units.

OVERVIEW

Predicting small group performance from the characteristics of individuals and groups is a complex and multidimensional process. The purpose of this review is to provide a broad survey of the results that have been found in both the civilian and the military literature that might have an application to small units engaged in combat arms. We have attempted as extensive a coverage of the military literature as

practical, as the relevance of these studies is obvious, but the vastness of the civilian literature mandated a focus as to units studied and topics addressed.

Our approach to civilian studies has been to cover the field of group performance in general with a concentration on one subfield: the psychology of team sports. This was done because the motivational and task characteristics of military and sports performance have strong similarities. Many of Zander's (1978) characteristics of athletic teams apply to military units as well: they perform in public, act as proxies for a larger public who are emotionally involved and who demand victory, are subject to public criticism and shame if they fail, are trained to operate with well-defined rules, and must spontaneously react appropriately to unexpected events. Indeed, leaders in each field freely employ terminology from the other. Both military and team sports tasks may be characterized as addressing highly competitive situations involving winning vs. losing as the very reason for the formation of the unit.

We should note, however, that examining performance of sports teams introduces methodological problems of generalization. Sports teams are voluntary bodies with a high degree of self-selectivity, and any conclusions based on behavioral observations of such bodies must be tempered by the fact that team members may be atypical of the general population on the behaviors measured and their underlying causes. While the modern Army is also a voluntary organization, whose members may not be representative of the population at large, there have not been any empirical studies showing that Army volunteers and sports team members are similar enough subpopulations to conclusively demonstrate the validity of generalizing from one to the other.

Nonetheless, we concentrated our literature search on predictors of performance of military and sports units, relying on summaries of the state of the art in large part for research on other units. We included specific studies from the general research literature either to illustrate the way the research community approached the topic at hand or if the studies made major theoretical or practical statements.

The topics considered here have emerged from the reviewed literature. First, we discuss the problem of units of measurement and define the type of unit we are investigating. In so doing, we distinguish among types of units and characteristics of groups. Then, we address the problems of defining group tasks and performance measures. These definitions examined, we turn next to an investigation of predictors of group performance. In this investigation, we first address general knowledge, and then turn to any findings directly applicable to military units.

A theoretical model that guided the organization of this review is Living Systems Theory (Miller, 1978). This theory is an extensive model of hierarchically organized living systems, of which a military structure is an obvious example. In Living Systems Theory, the behavior of units at different levels of organization (soldier, squad, platoon, . . . , division) is analyzed in a framework that postulates that any living system engages in two types of processes—those dealing with the physical world of matter and energy and those dealing with the symbolic world of information processing. These two worlds are further broken down into functions that have common representation in any level of organization, such as communication, transportation, or decisionmaking; these functions then become the targets of study.

While the potential usefulness of this approach is evident, there unfortunately appeared to be no explicit attempts to employ Living Systems Theory in our examinations of military unit performance. Ruscoe (1982) documents the Army's considerable interest in the model, but the research he reports is currently restricted to the battalion level of organization. Ruscoe's analyses, although not immediately pertinent to the review at hand, provide a model for the diagnosis of problems that can impede the effectiveness of unit functioning. Problems of appropriate measurement, treatment of data, and decoding of results are all illustrated from the Living Systems viewpoint, and some findings are directly translatable into recommendations for changes in procedure on the battalion level. Ruscoe concludes that the Living Systems approach is by itself insufficient to treat the empirical realities of U.S. Army organization, but does provide a framework on which to build a workable model.

Other critics of research on small group performance have noted features of the problem that are consistent with a Living Systems approach. MacCrimmon (1980) points out that tasks could be defined on their degree of complexity, the amount of uncertainty in the environment, and the conflict inherent in the situation. Each of these dimensions is applicable for analysis at different levels of organization, and the appropriate decisionmaking steps and criteria for good performance are not level-dependent. MacCrimmon's approach is theoretical rather than empirical or practical, but he does offer some practical applications backed by informal case histories. Roberts (1980) has noted the general eclecticism in the study of group performance research and calls for a more unified approach. Many of her recommendations are entirely consistent with the systems approach. These include:

- Examining functioning groups in their natural environment
- Studying group processes over time
- Studying transformation processes
- Regarding input/output and communications links as major process variables
- Focusing on the group level of analysis

Our analyses below will often address these points.

Our investigation identified five general categories of predictors of group performance:

1. Individual characteristics (general ability, task proficiency, and personality characteristics)
2. Leadership (ability of the leader, personality, and leadership behavior)
3. Group structural composition, or the *mix* of individual characteristics (general ability, task proficiency, personality, and cognitive style)
4. Group processes (cohesiveness, attraction)
5. Training techniques (feedback vs. no feedback, and feedback about group vs. individual performance)

The amount of detail given to each of these categories is related to the likelihood of application in the military setting. For example, a pressing question in the military is how to compose groups—whether those groups are crews, platoons, or companies—for optimal performance. Since the military has considerable control over how units are assembled, and may profitably use the results of research comparing different group compositions, considerable detail is provided for the research on group composition. Less detail is given to the research on leadership because this variable is not easily manipulated, either by altering the behavior of individuals in leadership capacity or the predominant leadership styles in established organizations. Furthermore, some characteristics that have been shown in civilian studies to relate to group performance are omitted here because they have little or no relevance to military settings, including, for example, whether leaders are appointed or elected, the amount of information given to group members regarding the group goal, and the communication structure (e.g., hierarchical lines of communication vs. fully interlocking networks).

II. DEFINITIONS

UNITS OF ANALYSIS

In a hierarchically arranged organization such as the Army, a vital issue that precedes any evaluation of performance is the choice of a unit of analysis. At times, the choice of unit of analysis might be self-evident from the task at hand. But at other times, this choice might be not at all clear, yet critical to the problem being faced. For example, in deciding who has won a battle between opposing armies, the unit of analysis is the entire force of each side; individual vs. individual combat or even wing vs. wing encounters are not of interest except as they define the outcome of the whole. Similarly, in marksmanship training, the score of the individual soldier is the appropriate unit of analysis; aggregate platoon or company scores might be helpful for other reasons, but are not useful in evaluating the accuracy of a particular rifleman. However, in group training exercises, the appropriate unit of measurement is not clear. If a rifle company is assigned the training exercise of taking a specified objective, is the appropriate unit the individual rifleman, a platoon, or the entire company? At what level is feedback best provided to the company so as to improve performance?

For many tasks, and particularly for combat tasks, it has become apparent that the individual soldier is often an inappropriate unit of analysis. Instead, especially for tasks which require coordination and cooperation among soldiers and for which success is measured for the working group as a whole rather than for its constituent individuals, the appropriate unit of analysis is a crew numbering from two to ten individuals. This small group will be the "unit" of the following review of the social psychological literature on the predictors and concomitants of unit performance.

DEFINING A UNIT

The task order for the present project mandates a study of unit performance, where that term is taken to mean the aggregate behavior of personnel within a unit. This excludes an assessment of equipment, weapons, or other logistical characteristics associated with units.

Although the size of the unit is not defined, it is implicitly assumed to be the smallest coherent aggregate of individual personnel.

Dyer et al. (1980) approached the problem of defining Army teams by surveying 11 of the 14 branches of the Army.[1] Experts in each of these branches were asked to identify all teams within their branch and characterize those types of teams in terms of their size, MOS (Military Occupational Specialty) of members, range of ranks of members, equipment, activities, and whether the team followed established (well-defined) or emergent (reactive to the environment) practices in its work.

A team for Dyer et al.'s purpose was defined as a small group, from 2 to 11 individuals (although some teams had as many as 40 members), whose roles were formally defined and whose tasks required at least some interdependence. Care was taken so that no "team" was defined as a combination of units each of which was itself a team. A total of 1248 species of team were defined by this procedure, which were collapsed analytically into 255 distinct types, which in turn could be fit into one of four global categories:

- Small homogeneous teams led by enlisted men
- Medium-sized homogeneous teams led by enlisted men
- Medium-sized homogeneous teams led by senior enlisted men or junior officers
- Large heterogeneous teams led by officers

Homogeneity and heterogeneity were defined by the number of distinct MOSs in the team and by the range in rank of the members. Having identified those teams through TRADOC[2] experts, Dyer et al. then surveyed 140 different units throughout FORSCOM to identify training and practical needs and problems of teams; some of these data are reported below.

Hall and Rizzo (1975), in a study of tactical team training for the U.S. Navy, followed a traditional distinction between teams and small groups. Teams are characterized as relatively well organized, highly structured, and with well-defined formal operating procedures. Members have assignments so that the participation of any one person can be anticipated by the other members of the team. There is generally some specialization so that subunits of members may be defined such that member duties across subunits do not overlap to any great extent. By contrast, small groups are more diffuse, have loose communication networks, and depend on the quality of independent indi-

[1]Infantry, Corps of Engineers, Quartermaster Corps, Air Defense Artillery, Field Artillery, Armor, Ordnance Corps, Signal Corps, Chemical Corps, Military Police, and Transportation Corps.

[2]See Acronyms, p. xv.

vidual contributions to the task. Hall and Rizzo's analysis fixed four characteristics of a Navy tactical team:

- It is goal- or mission-oriented. That is, there is a specific objective for the team to achieve.
- It has a formal structure. For military teams, this structure is hierarchical in nature.
- Members have assigned roles and functions.
- Interaction is required among team members.

This definition largely coincides with Dyer et al. (1980) and provides us with a consensus definition of a unit as the smallest interacting collection of individuals that has a functional identity.

DEFINING PERFORMANCE

On the surface, the definition of group performance is relatively simple: winning is better than tying, and tying is better than losing. In a sports competition, for example, over a season of matches, the more wins, the better the team has performed. Team training may be validated by performance; if a coach's techniques produce winning teams, his job is secure, but if his techniques falter, he is replaced. Military team performance is a different matter, though, for in recent years, there has been mostly training, and little battlefield testing to determine group performance. That this is a major military problem is well recognized (Hagan, 1981; Madden, 1981); therefore, much effort has been expended to construct exercises for military teams that provide measures that have face, content, or construct validity for the battlefield tasks that might eventually have to be performed.

Ryan and Yates (1977) assessed the face validity of Operational Readiness Training Tests (ORTTs) by asking the soldiers tested whether they felt the instrument was realistic and reflective of their performance. Results were generally positive for this behavior-based system; most recommendations about how to improve the ORTT program were in the direction of making it more realistic with respect to how the enemy might behave in combat. However, without some form of control for the type of instrument whose validity is being assessed, this positive result could be due to general cooperativeness on the part of respondents or other similar methodological artifacts.

Grunzke (1978) reported on an automated flight training system performance measurement package for the Air Force, and found that out of 28 dependent measures on the scoring format, only three variables discriminated between student and operational air crews, and for two of those measures, the students had a superior performance. He

concludes that the package has a potential to provide objectively scored performance measurement for crew performance and to provide an information feedback tool for air crew training, but more research is needed before this potential can be realized.

Obermayer and Vreuls (1974) and Obermayer et al. (1974) have developed a detailed crew performance measurement system for the Air Force, in which each phase of a combat flight is broken down into small behavioral segments for each crew member. The behavioral segments are then evaluated as correct or incorrect. For purposes of assessment of crew performance, summary measures may be constructed, whereas for training feedback purposes, the individual behavioral evaluations are available to the crew members. Such a system clearly depends on a judicious breakdown of behavior and the development of a system of feedback that does not overload the cognitive capacities of the crewmen in training. O'Brien et al. (1979) have provided a similar breakdown for assessing tank crew performance on the M60A1 tank. As before, behaviors of the individual tank crew members are assessed at a micro level in a way that can be reliably measured by experienced raters. The technique has the distinct advantage of requiring little subjective estimation, but is validated on its face rather than in comparison with any battle-tested indicators. If the constructors' theories of what constitutes good performance are correct, then the test exercise scores are valid and useful; if not, then it is not clear what the test is measuring.

Turney and Cohen (1981) and Turney et al. (1981) have attempted to define good Navy team performance by surveying the literature of good information transfer skills and developing from this survey indicators of coordination skill. Their review led to the conclusion that the team skill of coordination, whether arrived at because of superior team member characteristics or because of how the group was structured, was the important determinant of team achievement for a variety of team tasks. They have some recommendations of what is good and what is bad performance in this regard, but have not yet constructed evaluation instruments specific to any particular Navy task.

Several investigators have advocated detailed objective scoring of task segments in training evaluations. Havron et al. (1979) argue that engagement simulation techniques are not only superior as training techniques (see below), but also provide more objective evaluation criteria for team performance. Evaluators in engagement simulation exercises classify in detail the various tasks performed by the units and provide numerical, objective ratings instead of more global summary evaluations that arise out of earlier training processes. Similarly, Knerr et al.(1979) specify such variables as casualty exchange ratios,

mission accomplished scores, and extensive usage of process measures, especially employing computerized and other highly technological training tools (e.g., MILES) to provide more objective evaluations of training exercises. They argue that the present databases are not being utilized as fully as they might be, and more data from individual behavioral segments is needed. Madden (1981) complains that there is a lack of coordination between evaluation exercises and the training programs, which leads to poor performance because units are not trained to do what is required of them. He argues for a more integrated development of training and evaluation systems instead of the present incremental pace of change.

Hagan (1981) summarized thinking about the problem of measuring team performance by noting that evaluation is intimately tied up with training. As units are trained for particular tasks, so must evaluation processes be tied to those tasks. Then dissatisfaction with the evaluation instruments must lead back to the training system from which they arose and cause changes in that system, which in turn lead to changes in evaluation procedures. He discusses how this has been manifested in the development of ARTEP as a training, feedback, and evaluation device, illustrating both its successes and failures.

All of the articles cited above have emphasized the importance of objectively defined measures of performance as opposed to subjective global evaluations by an expert or superior officer. This emphasis is consistent with a well-established finding within psychology that "objective" predictors do a better job than "subjective" or "clinical" predictors for a variety of areas ranging from prognosis of mentally ill patients to predicting performance in psychology doctoral programs. Even acknowledged experts who "feel" that they can best *know* a person through an interview are outperformed by relatively simple linear regression models based on objective predictors.

Perhaps one reason for the superiority of objective ratings is that subjective evaluations are biased by impressions of effort, rather than being pure measures of achievement. Indications that this may be the case come from studies of the attribution of success and failure in sports competition. Iso-Ahola (1976) shows that sports teams were differentially rewarded or punished more on the perceived effort expended than on their actual outcome. High effort was rewarded no matter what outcome obtained or how capable the team was perceived as being. On the other hand, low effort was punished, especially when a high-ability team barely won, or worse, lost.[3] Bird and Brame (1978)

[3]Low-ability teams were not punished as severely for lack of effort; presumably when faced with futility, giving up is permissible.

indicate that members of losing teams may separate their own evaluation from that of the group, adopting an "I'm O.K., but the team's so-so" attitude in which their own effort is seen as greater than the collective effort of the team. That is, winning teams try harder, but losing players try harder than their teams. Because of this additional effort, their personal evaluations of outcome are not as bleak as those for the team.

No corresponding studies of the commingling of effort and performance were found for studies of military performance, but the folklore on the value of effort, plus decades of evidence in other arenas, suggests that the contemporary emphasis on objective performance measures is well-placed. Therefore, this review, and our suggestions for future research, will focus on objective performance measures.

Once performance measures are obtained, the question of how to treat them appropriately arises. In particular, attention must be paid to the *reliability* of a performance measure (how consistent is the measure) and the *validity* of the measure (to what extent does the performance measure assess the qualities it purports to assess). These are technical questions which are applied when appropriate to the individual studies. In addition, we have written a prescriptive essay to guide future research efforts to reliable and valid unit performance measures; this essay is an Appendix to the present review.

DEFINING GROUP TASKS

A study was eligible for inclusion in this review if it reported on a group that produced a group score. Moreover, all group members had to know that they were members of the group and supposed to be working toward a group goal. This criterion ruled out studies in which group members were unaware that the group was being evaluated on its achievement of a collective task.

Even restricting the research to tasks with known group goals, the variety of tasks is large. The tasks can be grouped in four categories according to the amount of interdependence required among group members, whether members performed the same task or different subtasks, and whether the activities of each member were specified by the task requirements.

In the first category are tasks that required no interaction among group members. These tasks qualify as group tasks only because group members worked toward the same goal. In one example, group members built models of molecules individually; the group's score was based on the total number built (Hewett et al., 1974). In another

study, group members sat at separate consoles and pressed buttons in response to certain light stimuli; the group earned a point when a prespecified number of group members reacted within a certain time interval (Zajonc, 1962). Marksmanship scores that are a sum of individual performance scores, or armor company scores that are a sum of individual unit proficiencies are other examples of this kind of task.

Second are tasks that required division of labor. These tasks required each group member to work on a different subtask, but the subtasks formed one group product which was then evaluated. One example is building a chart using data given to the group, wherein group members were assigned different sections of the chart (O'Brien and Owens, 1969). Another is the surveying task of Terborg, Castore, and DeNinno (1976), in which one member of a triad worked the plumb line, another operated the transit, and the third wrote the results. In these tasks, group members interacted only to combine the products or results of the separate subtasks. Within the military, tasks of this type typically are formulated on large organizational levels of analysis, such as a movement of infantry forces after an artillery barrage in a single operation. The infantry's success is conditional on the quality of the artillery's performance, but the two do not interact.

In the third category are tasks that required interdependence by all group members. These tasks could not be performed without the coordination of all group members. A good example is the motor maze task used in Gill's (1979) study, in which members of a dyad operated different controls that tilted the maze board. In another study, different group members operated different controls in a model railroad (Ghiselli and Lodahl, 1958). This category is possibly the most relevant to military units such as tank crews, in which crew members must coordinate their efforts to achieve success. Within the civilian literature most examples come from sports teams, such as basketball, volleyball, or ice hockey units.

Finally are tasks that called for collaboration by all group members but in which groups were free to pool resources in any way they chose. Most of the studies used this kind of task. Several examples include writing Army recruiting letters (O'Brien and Owens, 1969), intellectual problem-solving (Triandis, Hall, and Ewen, 1965), creative writing (Sorenson, 1973), group-taking of intelligence tests (Laughlin and Branch, 1972), decisionmaking (Lampkin, 1972), brainstorming (Bouchard, 1972), and designing computer systems (Hill, 1975). Theoretically, all group members may participate if they choose, but the task can often be completed by one person or by a subset of group members.

The major division among these four categories of tasks is between the second and third types, and corresponds to the distinction made in the team sports literature (Bird, 1978; Gruber and Gray, 1981; Landers et al., 1981; Landers and Luschen, 1974; Widmeyer and Martens, 1978) between *coacting* and *interacting* tasks. An interactive task is one which requires group members to actually work together and coordinate their efforts in order to reach a group goal. For example, although a single player actually crosses the goal line to score in football, this act is largely impossible without the coordinated effort of the other 10 team members. In contrast to this is a coacting task, where the members have a common goal (team victory), but where their contributions are more individual and less differentiated in nature. For example, members of a golf team or marksmanship team coact. The distinction is not a discrete one; many tasks are intermediate on this continuum. For example, on a baseball team, batting is a coactive task, but fielding is largely interactive. Most military tasks, especially in combat units, may be classified as interacting; Hall and Rizzo (1975) specifically require interaction, although Dyer et al. (1980) do not in their respective definitions of team tasks. Certainly the tasks of members of an armor team, with the division of labor into commanding, driving, gunnery, and loading, is interactive, but riflery, albeit with some specialization, has elements of coactiveness as it can depend to a large extent on the individual independent performance of its members.

The abundance of loosely defined collaborative tasks in the literature on predicting group performance makes it difficult to interpret results unless the degree of interactiveness of the task is known. Because the relations between group performance and either individual characteristics (e.g., member ability and personality) or group characteristics (e.g., group composition on the basis of proficiency or attitudes) and group performance depend on the task requirements for interdependence among group members, results are often ambiguous. In discussing the results of research on predictors of group performance, therefore, considerable detail about the group tasks is given to identify the degree of interactiveness as much as possible. In general, we will concentrate on interacting tasks which are characterized by aiming at a common goal to which all members aspire, by a division of labor among members, and by the requirement of coordinated effort among members.

III. PREDICTING GROUP PERFORMANCE FROM CHARACTERISTICS OF INDIVIDUALS

We begin at the most microscopic level of analysis, looking at the characteristics of individuals that are predictive of the performance of groups containing those individuals. The individual characteristics used to predict group performance include general ability, task proficiency, and personality characteristics. In determining best predictors, studies have used a variety of data reduction techniques to obtain an ability score to correlate with group performance measures. Among the scores used have been the ability or proficiency of the most-able group member, the least-able member, the mean ability of members, and the ability of a centrally located member.

Each of these measures is an index of individual characteristics in the group, as opposed to an index of the mix of characteristics over unit members. Our immediate interest is in the contribution of individuals such as the most-able and least-able member, with some attention paid to the contribution of the "average" member, as measured by the mean or sum of individual scores. The mix of characteristics will be considered in Sec. V, including a closer look at group means and an examination of homogeneous vs. heterogeneous groups. Although correlations between predictors and performance were often fairly high, the specific patterns of correlations for the various ability indices differ across studies.

GENERAL ABILITY

Although we would expect the general ability of group members to be positively related to the performance of the group (see, e.g., Hare, 1976; Bass, 1980), the strength of the relationship seems to depend on the characteristics of the group task. O'Brien and Owens (1969) cite two studies that distinguished among different kinds of task requirements. These studies examined the correlation between member ability and group performance for both "collaborative" and "coordinated" organizational structures. A task was collaborative if all group members worked together. In a coordinated task, group members were instructed to work on separate subtasks. In our own terminology, these would be interactive and coactive tasks, respectively. One study used Australian army soldiers assembled in four-person groups to write

recruiting letters (interactively) and to construct charts showing the results of examinations given at military schools during earlier years (coactively). The second study was a laboratory task, in which three-person groups wrote stories from Thematic Apperception Test (TAT) pictures. In the interactive condition, all group members worked together on each story, whereas in the coactive condition, each member worked on a story for 20 minutes and passed his story to the next person. In a third, mixed condition, group members worked together for 15 minutes and then rotated as in the coactive condition. The ability measure used in the first study was the Army General Classification Test (GCT), and in the second study was the American College of Testing subtest on English.

O'Brien and Owens correlated group performance on each task with different group ability measures, including the sum of abilities within the group,[1] the ability of the least-able member, the ability of most-able member, and the ability of the group leader (the leader in the Army study was the member with highest rank; the leader in the laboratory study was appointed by the experimenters). In the Army coactive task, the first three ability indices (sum, lowest, and highest) were significantly associated with group performance (correlations ranged from 0.48 to 0.58). In the laboratory coactive and mixed tasks, only the sum and lowest scores were related to group performance (correlations ranged from 0.49 to 0.56). Over all coactive tasks, then, general ability accounted for between 23 and 33 percent of the variation in group performance scores. For none of the interactive tasks in either study was ability significantly related to group performance. A possible explanation for this finding lies in the way unit members may organize themselves in coactive and interactive tasks. In the coacting tasks, since all group members had to contribute to the task, group performance was in part determined by the abilities of all members, including the least-able member; no one person could cause the quality of the group product to be high. By contrast, in the interacting tasks, groups were free to combine their talents any way they chose. Although the investigators did not discuss this possibility, perhaps some groups in the collaborative condition depended on the efforts of the ablest member if that person was very superior, whereas others used the collaborative efforts of all group members. If different groups each selected their most efficient strategies to perform their tasks, then the observed nonsignificant correlation between ability and group performance would result.[2]

---

[1]The sum of abilities and mean ability are equivalent statistics if the group size is held constant, as is the case in most experimental studies.

[2]In the language of Living Systems Theory, there is equifinality among strategies; put more commonly, there is more than one way to skin a cat.

Turning to military studies, there have been several attempts to assess how individual ability affects unit performance. These studies are all similar in that they

1. Measure the contribution of specified individuals (e.g., tank commander) on unit (e.g., tank) performance,
2. Do not measure the joint effectiveness of two or more members, much less how ability mixes affect joint performance, and
3. Differ from each other largely on the choice of what is used to predict performance. The most popular indicator of ability used to assess performance has been the Armed Services Vocational Aptitude Battery (ASVAB), administered routinely to all recruits for ten years.

Black (1980) used a composite of ASVAB scores called CO (for combat potential) to attempt to predict the performance of gunner/loader and driver performance during training and, later, exercise performance of tank crews. Performance quality was measured by summing dichotomous correct/incorrect evaluations of the subtasks involved in taking a tank through a field exercise. Black found that the CO measure predicted success for gunner/loaders and drivers while they were in training, but that when more experienced men were tested, there were no differences. A likely explanation of this finding is that CO, which has a number of cognitive components, is an indicator of the speed of learning. During training, unit members with high CO scores will learn faster, and therefore perform better, than members with low CO scores. However, with experience, most unit members reach a competency level of acceptable performance (as emphasized by the dichotomous nature of the performance measure components), and the differences vanish. Were the performance measures to be more finely graded, the differences due to CO might be detectable.

Maitland (1980) and Maitland, Eaton, and Neff (1980) also used the ASVAB to predict the performance of tank gunners and drivers. Maitland (1980) used the entire battery of ASVAB scores to predict performance defined as successfully accomplishing subtasks involved in firing the main gun, driving the tank, and other MOS-specific tasks required in tank exercises. Separate multiple regressions were done for 130 driver trainees and 205 gunner/loader trainees. For drivers, the ASVAB measures of numerical operations, arithmetic reasoning, automotive information, and electronics were retained, whereas for gunner trainees, word knowledge, mathematics knowledge, and mechanical comprehension were used. These sets are a partial overlap with

Black's (1980) CO score. For each crew specialty, a unit weighting scheme was used where the individual's standardized scores (z-scores) for each of the items in the equation were summed to produce a single composite. "Validities" (meaning multiple correlations) were in the high 20s for both prediction equations, which indicates statistically significant but relatively moderate associations. Maitland et al. (1980) report on retests of the predictors over time. The predictors were valid for a retest of trainees soon after the first testing period, and were similarly valid for a group of former trainees tested at a later stage of their training. But a cross-validation based on experienced crewmen showed a considerable weakening of the predictors' validity, a finding that is similar to that found by Black (1980) as reported earlier, and probably attributable to the same phenomena.

Eaton (1978a) administered a battery of paper and pencil instruments to members of 51 tank crews assembled for annual qualification tests. Performance scores for tank commanders and gunners were based on objective Table VIII[3] test results, but for drivers were based on subjective rankings by platoon leaders. Each dependent variable from the test was used separately in a multiple regression test. Eaton explicitly acknowledges the problems arising from the low sample size to predictor ratios he employed. Overall, some of the predictors were statistically significant, but even for these it is not clear how the measures contribute to unit performance variance, or what they imply for improved training or recruitment measures. For tank commanders, successful predictors included tests of object completion, pattern recognition, and mechanical abilities; these combined to predict half of the variance in the number of successful engagements in a Table VIII run; other regressions were not significant. For gunners, although no multiple regressions were statistically significant, Eaton indicates that visual recognition is related to first-round hits and that lateral perception and attention to detail appear to predict the amount of time spent per engagement. For drivers, nothing appeared to predict performance well.

Eaton, Bessemer, and Kristiansen (1979) used ASVAB scores plus paper and pencil measures to predict driving and gunnery performance of armor crews. Their objective was to obtain measures that indicate whether a recruit will best be trained as a gunner/loader or a driver.[4] In the first phase of the project, multiple regression techniques were used to predict driving and gunnery performance of recruits. Several variable sets were found that were good predictors. However, in a

[3]Tables VI and VIII are Army field tests for tank crews.

[4]These tasks have recently been assigned separate MOS designations.

cross-validation on a second sample, no replication of the findings were found except that the automotive comprehension score of the ASVAB was related to driving performance. Finally, using experienced crews in Germany taking a Table VIII annual exercise, none of the predictors successfully accounted for performance. Eaton et al. (1979) conclude that their present paper and pencil tests do not predict performance, and that other measures must be sought.

In summary, then, a sharp contrast exists between civilian and combat military tasks when performance is predicted from general ability. This contrast is partly due to a more methodologically sophisticated approach on the civilian side, but could also be attributable to the nature of the tasks employed in the research studies.

The military tasks surveyed have been field performance measures, involving physical and spatial skills as well as cognitive ones. Additionally, these tasks had high saliency for the soldiers, whose evaluations (and hence salaries, promotions, and possibly even careers) were on the line. In contrast, the civilian tasks were self-defined experiments of no intrinsic value to unit members, and involved only cognitive skills. Even so, the O'Brien and Owens (1969) studies suggest that when *all* group members contribute to the task, group performance will depend on the ability of all of the group members, including the least-able member. However, when group members are not instructed how to combine their resources, no prediction can be made. The military tasks investigated are all interactive, but we do not have any indication of any long-term contribution of general ability to performance for the tasks examined. We recommend that studies examining performance as a function of combined crew member abilities be undertaken, and that the range of military tasks studied be expanded.

INDIVIDUAL TASK PROFICIENCY PRIOR TO TEAMWORK

A number of studies have used proficiency of individual group members on the task, rather than measures of general ability, as the predictor of group performance. The problem with such studies is that the importance of individual performance may be quite task-specific, so that no generalization over tasks is possible and separate assessments must be made for any new task. The tasks used in civilian studies include maze problems (Gill, 1979; Rohde, 1958; Meister, 1976), the Purdue Pegboard Test (Comrey, 1953), jigsaw puzzles (Wiest, Porter, and Ghiselli, 1961), crossword puzzles (Comrey and Staats, 1955), and light-switching (Egerman, 1966; Klaus and Glaser, 1965). The motor

maze task used by Gill (1979, p. 115) serves as a good example of an interactive task:

> The object of the task is to move a steel ball through the maze as quickly as possible while avoiding numerous *cul-de-sacs*. To move the ball through the maze the person operates two handles located on adjacent sides of the maze that tilt the maze board forward and backward and side-to-side, respectively. . . .
>
> For the [interactive] group task each of the two group members used one handle of a single maze. The task required maximum collaboration because the two handles must be operated together to successfully negotiate the maze.

The performance measure was the time taken to complete the maze. In the Purdue Pegboard Test, dyads assembled towers of pegs, washers, and collars: member A inserted the peg in a hole, member B placed a washer over the peg, member A placed a collar over the washer, and member B placed a second washer over the collar. Group members alternated assignments on each assembly. The number of completed assemblies was the performance measure. In the light-switching task, members of a dyad had to press a light switch for either two seconds or four seconds for each of several stimulus light patterns. Although each group member had a separate control panel, the team did not score a point until both group members gave an accurate response. The group performance measure was the number of points earned in a specified amount of time.

In all of these studies, group members learned the task as individuals before working as a team. Individual proficiency was determined during a test period following training.

The results showed three seemingly contradictory patterns. First, Gill (1979), using the motor maze task, found that the pregroup proficiency of the slower member of the pair significantly predicted group performance (correlations across experiments ranged from 0.50 to 0.71), but the pregroup proficiency of the faster member did not predict group performance (correlations were near zero). Second, Comrey (1953, pegboard test) and Wiest et al. (1961, jigsaw puzzles) and Comrey and Staats (1955, crossword puzzles) reported significant correlations between *all* measures of proficiency, including the pregroup score of the most proficient member, the pregroup score of the least proficient member, and the sum of the scores in a group (correlations ranged from 0.56 to 0.79). Third, Rohde (1958, maze task) found that the proficiency of the most-able member and the sum of the proficiencies of the three members of the group predicted group performance

(correlations were 0.63), but the proficiency of the least-able member did not predict group performance.

Klaus and Glaser (1965) formed homogeneous groups (high, medium, low) using the level of proficiency of individual members on a light-switching task, as in the studies described above. They also used speed of learning (fast, slow) as a factor in their design. Pregroup level of proficiency predicted group performance, but speed of learning did not. Therefore, what mattered was an individual's final level of attainment before group work, not the time taken to reach that level. This finding would also explain the generally poor predictability of general ability in the armor studies. Tank crews are not randomly selected individuals, but are instead screened before training so that they constitute a set of people believed able to attain proficiency in the tasks they are trained to perform. They thus form an attenuated group with a restricted range of ability. Within that range, ability does not predict performance, even though the relationship might be important over the entire range of ability.

Two studies from sports psychology have examined team performance as a function of the abilities of individual team members. Widmeyer, Loy, and Roberts (1979) examined the contribution of the ability of each individual to the success of doubles tennis teams. Only nine players were considered, and individual ability was assessed by raters, which qualify the results. But the nine players formed 33 (out of a logically possible 36) different teams, and many matches contributed to the data. Through multiple regressions on the dichotomous win/lose dependent variable, the combined ability of both players accounted for 29 percent of the variance, indicating that it was a major, but not decisive factor.

Jones (1974) examined archival data on professional sports teams to assess the contribution of ability to team outcome for tennis, football, baseball, and basketball. His measures differed widely over sports. For tennis, he used the United States Lawn Tennis Association rankings of singles players to predict the same organization's rankings of doubles teams, assuming the ranking was intervally scaled. For football, he abandoned individual measurement and calculated separately the quality of the offensive and defensive units of National Football League teams on the basis of points scored and points given up to predict won/lost records. For baseball, a technique similar to football using pitching earned run average as the defensive measure and team batting average as the offensive measure was used to assess major league teams. Finally, in basketball, the productivity of the best five men on National Basketball Association teams was employed. For each of the four sports examined, a linear sum of ability was a good predictor of outcome; the proportion of variance accounted for ranged from 36 to 90

percent. The relationship was strongest for baseball and weakest for basketball.

One study of tank crew performance employed job-specific predictors of ability. Eaton, Johnson, and Black (1980) attempted to improve predictability of performance by moving from paper and pencil predictors to job samples. Two tasks were employed as predictors, the first a tracing procedure where subjects had to use vertical and horizontal controls on a terminal to move a cursor through a figure on the screen without crossing the figure's boundaries. Two figures, a diamond and a circle, were originally used, but tracing only the diamond (much the easier task) eventually was retained as the indicator of performance. Two measures were employed, the time to trace the diamond and the number of errors (cursor outside of the boundary) produced. The second was a sensing task, where subjects were requested to locate where a round had landed on a picture simulating a firing of the main gun. In a sample of 47 experienced gunners/loaders, it was found that the error scores on the diamond tracing task successfully predicted performance on a Table VI gunnery exercise. Eaton et al. (1980) then replicated these findings using 24 gunner trainees who had recently graduated from the training course. Again, errors in diamond tracing predicted gunnery scores of the tank, including total hits, first-round hits, second-round hits, and moving target hits. In this replication, tank drivers were also tested, and it was found (contrary to expectations) that they did not fare worse on these gunnery predictors than did gunners.[5] Finally, a third sample of 160 trainees was employed, broken down into beginning or mid-training experience and whether they got round-by-round feedback on the training task. ASVAB scores were also used as predictors. It was found that with training, subjects performed better on the tracing and sensing tasks, but there were no differences on Table VI performance attributable to any of the job sample or ASVAB predictors.

At first glance, the different patterns of results appear to be contradictory. However, the O'Brien and Owens (1969) work reviewed in the previous section suggests that the nature of the task can at least partially explain the inconsistent results. The studies in which the proficiency of the least-able member significantly predicted group performance were coactive tasks, in which each member contributed independently of the other members. For such tasks, errors and inefficiencies are propagated through the task, with little chance of correction. On the other hand, the tasks showing significant relationships between the proficiency of the most-able member or composite ability

[5]Table VI performance was not regressed on these driver scores.

were interactive, such that individual errors could be compensated for by correct group performance, or interdependence such that more-able members could pitch in to help weak members. Similar analyses have been done in research involving decision rules for group bodies such as juries, where rules are promulgated that maximize the likelihood of desired outcomes, such as voting for the true state of affairs. Such studies, however, fall outside of our performance purview because of their emphasis on purely cognitive tasks; they offer little in the way of advice for improving unit performance in the sense defined for the present project.

Although the specific patterns of correlations may have differed somewhat across studies, a general finding emerges from most of them. When the proficiencies of the most-able and least-able group members were used as predictors of group performance in multiple regression equations, the multiple correlations ranged from 0.54 to 0.72, showing that between 29 and 52 percent of the variation in group performance could be explained by individual proficiency. This is, by social scientific standards, a sizable proportion of the variation, and merits policy recognition. The remaining variance is to be explained by other factors, including other characteristics of individual members (e.g., personality characteristics), combinations of characteristics of group members (group composition indices), and group processes.

PERSONALITY AND MOTIVATION

Much of the work relating personality factors to group productivity is rudimentary. For example, Maksimova (1973) reports that collective job efficiency in Soviet collectives is related to individual industriousness and responsibility and negatively related to authoritarianism, modesty, and shyness. However, all measures were subjective evaluations, and no reliability or validity criteria were reported. Although these findings might seem believable on their face, it is difficult to distinguish cause from effect in such subjective evaluations.

The most common method used to relate personality and group performance has been to calculate correlations, based on individual scores, between group performance and a large collection of personality scores from one or more personality inventories. In a typical analysis, members of each group are assigned their group's performance scores in addition to their own scores on a battery of personality tests. Two studies will be reviewed here to illustrate this type of research (see also Heslin, 1964). One of the earliest studies of this sort was conducted by Haythorn (1953). In Haythorn's study, four-person groups worked on

reasoning, mechanical assembly, and discussion tasks, and independent judges rated each group's productivity. Group overall productivity (across the three tasks) was correlated with members' scores on the Cattell Sixteen Personality Factor Questionnaire. Only two of the 11 personality factors reported by Haythorn were related to group productivity: emotional stability (0.48) and "Bohemianism vs. practical concernedness" (–0.61). Unfortunately, Haythorn did not present the correlations separately by task.

The other study (Bouchard, 1969) correlated the variables on the California Personality Inventory with group performance on brainstorming and a critical problem-solving task (how to maintain a high standard of education in the face of growing enrollment). On the brainstorming task, group performance was predicted by a group of variables representing interpersonal effectiveness (dominance, capacity for status, sociability, social presence, and self-acceptance), tolerance, and intellectual efficiency (correlations ranged from 0.30 to 0.52). On the problem-solving task, group performance was predicted by capacity for status and sociability (correlations ranged from 0.24 to 0.38).

Because so few of the correlations among personality variables and group performance measures were significant (two out of 11 in the Haythorn study; 16 out of 90 in the Bouchard study), they may be chance results. More importantly, the correlations are difficult to interpret due to the analytic method. Significant correlations based on individual scores provide little understanding of whether the characteristics of all group members or only a few need be taken into account. The overall correlations mask the relationships between group performance and the most extreme score in the group (high or low), the mean of the group, and the variation in the group.

A small number of studies have been more oriented to testing hypotheses about personality, motivation, and performance, and have focused on specific personality measures rather than taking a "shotgun" approach.

Cooper and Payne (1972) investigated the relationship of personality orientations and performance in football (soccer) teams. They obtained the cooperation of 17 of the 22 English First Division football clubs in 1965, and administered personality inventories to players, coaches, and managers. Their primary instrument assessed motivation in terms of self-, interaction-, and task-oriented motivation. Contrary to hypotheses, no global differences were found among teams on the basis of their league success; however, coaches of winning teams did have more of a task orientation than did coaches of losing teams. In general, backs (players in defensive positions) were less self-oriented than forwards or midfielders (attacking players), which might be

attributable to the skills required for high-caliber football. Configurations of players in winning teams showing a high self-orientation at the expense of task and interaction orientations were also found. These personality differences are suspect, though, as they might be attributable to situational differences based on the time of testing and player expectations. Testing was done in the spring, when only two or three winning teams were still in contention for the national championship; for the losing teams, task orientation may be inappropriate for coaches, who are thinking about next year and their jobs. Similarly, the self-orientation of players on winning teams may arise from the fact that they were being considered for placement on the national team, an honor that is regarded very highly among professional soccer players.[6] From this study, it is not possible to draw any firm conclusions relating motivational orientation and group performance.

Butler and Burr (1980) administered a large questionnaire to 914 male U.S. Navy enlisted personnel. Their primary goal was to demonstrate three types of locus-of-control personality orientations: internal, external/powerful other, and external/chance. Factors identified with these three orientations were found; however, these types did not relate well to any of a variety of health- and job-related performance measures in military environments. The only correlation over 0.30 in absolute value was one showing a relationship between health and an external/powerful other orientation. As no nonmilitary personnel were tested, whether these findings have special significance for military performance is not clear.

In the only study located relating personality style specifically to military performance, Roberts, Meeker, and Aller (1972) compared the action styles of Naval officers to their performance in a decisionmaking game. The officers were typed as to their view of causality in life as attributable to force or strategy. An instrument developed by Roberts, an anthropologist interested in classifying games in different cultures, was employed. The officers were classified as pure force, pure strategy, or mixed thinkers. For the decisionmaking game, groups of five to eight officers were randomly constructed, and their performance was related to the styles of the group. It was found that having multiple styles represented yielded a superior performance to any homogeneous group, including one of mixed thinkers.

Several studies have examined individual motivation and performance in the military. Bessinger (1974) administered a morale inventory to recruits after the first, second, fourth, and sixth week of basic

[6]Indeed, England, as host country, won the World Cup competition in the year following this study.

training, and compared this level of morale to several outcome measures, including the basic physical fitness test, a comprehensive performance test, and instances of absence without leave (AWOL). Analyses were not on the individual soldier level, but rather on the company level for 18 training companies. No correlation was above 0.30 in absolute value for any comparison made, indicating that morale, as measured on the group level, did not affect company performance. Given the nature of the measures used, a unit of analysis of the individual soldier might have yielded different and more informative results.

Bauer, Stout, and Holz (1976) developed a model of predictors of discipline problems in the Army by interviewing over 1500 people on U.S. Army bases in both CONUS and West Germany. Using modern nonmetric multidimensional analyses, they isolated three components of discipline in the Army: good performance, good appearance, and good conduct. These factors emerged for combat and support units, and the first two held for training units as well. They then employed multiple regression to find what characteristics of units predicted good discipline. Performance was found to be predicted by a solid esprit de corps, good leadership, and satisfaction with work role. This study, which is methodologically sound in terms of its sample size, appropriate use of statistical tools, and representativeness of subpopulations, provides insight into environmental factors affecting positive motivations toward performance; the next steps are to ascertain ways of providing those motivations, and to show how motivations lead to performance.

Eaton (1978c) attempted to find out what sort of incentives motivated members of tank crews to perform well. He created a questionnaire which was administered to 52 experienced armor crewmen, and obtained composite measures assessing personal recognition, tangible reward, intrinsic satisfaction with a job well done, and self-actualization motivation for tank crews. This was followed by administration of the questionnaire to 220 crewmen to measure the relative strengths of those motivations. For tank commanders, loaders, and drivers, but not for gunners, he found that recognition was the most dominant motive. Tangible rewards, contrary to expectation, were rated slightly negatively. In practice, however, a combination of tangible reward (days off plus cash) and recognition (public commendation) for high performance was shown to increase crew efforts. Eaton concludes that recognition is an effective motivation whose judicious use can probably improve unit performance.

SUMMARY

The literature relating individual characteristics with group performance shows substantial correlations between member ability and proficiency on the task and group performance. However, those substantial correlations did not obtain for studies of military performance. Even in the various civilian tasks, the specific relationships between member ability and proficiency and group performance differ with the proficiency of the most-able member, least-able member, or the average of the proficiency levels in the group having different predictive power for different tasks. When the task requires contributions by all group members, the proficiency of the least-able member and the sum of all proficiencies in the group are good predictors of group performance. When the task can be completed successfully using only the proficiency of the most-able member, the proficiency of other group members may not correlate with task success.

These findings, both positive and negative, suggest avenues of research that should prove fruitful in ascertaining the contribution of individual characteristics to unit performance:

- Military studies of general ability have not to date examined the relationship of ability of all of the crew members to performance, only the ability of different members singly as they relate to the performance of the unit. Such studies, using the group as the unit of analysis, should supplement extant studies.
- Performance in a group task is clearly affected by the interrelationships of individuals' tasks within the group. For interactive groups, there is a need for close examinations of the processes by which groups complete tasks in order to discover how the group's problem-solving strategy is a function of the ability composition of the group. If any policy regarding group composition is undertaken, it would best be accompanied by concomitant structuring of interactive tasks to promote efficacious problem-solving strategies.
- Investigations of individual proficiency in tasks as proficiency relates to group performance have studied individual proficiencies on a common task, whereas most applications involve crew members with different roles. Moreover, the extant studies lose sight of the fact that a task performed by groups of people might have quite different demands than the same task performed individually. To obtain a clearer view of the role of individual proficiency, studies which examine within-role proficiency as it relates to group performance are needed. Such studies would necessarily be methodologically complicated and

require careful analyses based on measurement frameworks, as discussed in the Appendix.

The literature presented only weak or inconclusive findings for the effects of personality characteristics of group members on performance. Although this finding is somewhat disappointing, it is not entirely unanticipated; indeed, there are personality theorists (e.g., Mischel, 1968; Skinner, 1975) who hold that the very concept of personality traits as broad predispositions to behavior is not tenable. These theorists would expect variations in correlations across studies as a function of group composition and task requirement, as we have observed, but would anticipate no relationship between trait measures and general task performance. The debate over the existence of traits notwithstanding, it appears reasonable to conclude that personality measures do not presently provide a good means of predicting unit performance.

Motivation was shown to be important, as might be anticipated. However, the various studies show that intuitive predictions of what are effective motivators may not be valid; for specific units, a study of what unit members wish to obtain from performance is useful in constructing an incentive structure that will motivate good performance.

IV. LEADERSHIP

The group leader is the most important individual in a group, and one whose particular characteristics are most likely to be related to group performance. The literature on leadership is voluminous and varied. Although the many studies relating leadership and group performance are often reviewed as a single category, they may address very different issues.

There have been several studies of military leadership, most of which are oriented toward lower-echelon leaders of teams rather than higher-level command personnel. One exception to this general rule, which is noteworthy because of its inherent interest and its methodological innovativeness, is a study by Simonton (1980) designed to examine the individual and situational determinants of victory and casualty ratios in major land battles. Simonton examined 326 major battles throughout history in an effort to find those factors that would enable him to predict which general won. His predictor variables included individual aspects of the two competing generals such as their years of experience, number of consecutive victories before the target battle, and age; they also included situational variables such as army size, home defense, divided command, and year the battle took place. His analytic approach was to use stepwise discriminant functions to predict the winning general. This is a procedure that takes into account the extensive dependencies among the variables and provides the most predictability in the fewest number of measures. Although this procedure might cause a theoretically oriented investigator to miss potential connecting constructs between elements of his theory, it is an excellent technique for the applications-oriented investigator who is primarily interested in useful prediction instruments. Simonton found that four variables enabled him to predict 71 percent of the time who would win the battle. These were the differences in years of experience between the competing generals, the difference in length of "winning streak" (consecutive encounters won) between the generals, the taking of the offensive (i.e., choosing when to begin an engagement), and having a divided command (e.g., allied nations each with its own general). The first three are individual variables, whereas the fourth is situational, reflecting perhaps the conventional wisdom that two heads are better than one. For predicting casualty ratios, the difference in cumulative victories between the generals, the advantage in army size, having a divided command, and year of the battle all were effective

predictors. With regard to the last measure, apparently ratios have been evening out as warfare becomes more mechanized and less hand-to-hand combat.

This study illustrates several methodological points about leadership and group performance. First, there is the question of performance itself: leadership is often evaluated in terms of its results, making the separation of leadership behaviors which should be promotive of group performance and the performance itself difficult. This issue is particularly exacerbated when leaders are evaluated by subjective ratings of observers or superiors. Then, the tendency to equate quality of leadership with quality of group performance is very pronounced. Second, the distinction needs to be made between leader behavior and characteristics of the leader. The former refers to behaviors performed during the act of leading. Leadership styles (e.g., democratic vs. autocratic or initiative-seeking vs. conservative) fall into this category. The latter refers to ability and personality characteristics of the leader. Studies examining characteristics of the leader as predictors of group performance sometimes are really concerned with leadership behavior; the distinction should be maintained, however, because the implications for improving performance vary depending on whether leaders are to be better selected or better trained. Therefore, we next review separately the literature on the ability of the leader, the personality of the leader, and leader behavior.

ABILITY OF THE LEADER

As pointed out above, ability of the leader refers here to the general or task-related ability of the leader, rather than to the behaviors of the leader when carrying out leadership activities. A program of research by O'Brien and colleagues (O'Brien, 1968; O'Brien and Harary, 1977; O'Brien and Ilgen, 1968; O'Brien and Owens, 1969) has investigated the relationship between ability of the leader and group performance. To illustrate this research, two of the studies will be described here.

O'Brien and Owens (1969) conducted two experiments: an Army study in which groups wrote a recruitment letter or constructed a chart, and a laboratory study in which subjects wrote stories from TAT pictures (the experimental procedures used in this study have been presented earlier). In the Army study, the leader was the group member with the highest military rank. In the laboratory study, the leader was appointed by the experimenters. General ability was defined on the basis of the Army General Classification Test (GCT) or American College of Testing scores in English. For no task in either

experiment was the correlation between the ability of the leader and group performance statistically significant. This is not particularly surprising, for several reasons. First, general ability may not correlate highly with performance on the task. Second, it is not clear that the leader in either experiment was given any special function or responsibility in the group.

The second study, by Kabanoff and O'Brien (1979), deals with both of the limitations of O'Brien and Owens' study described above: the generality of the ability measure and the function of the leader. Kabanoff and O'Brien used the creative ability of the leader to predict group performance on creative problems (e.g., improvement of toys, unusual uses of objects). Furthermore, group leaders were the only group members to receive instructions for the task. The results of the study were complex. Although there was a significant main effect for leader ability, showing that groups with high-ability leaders performed better than groups with low-ability leaders, leader ability interacted with the type of task structure. In the recruitment letter task, group members worked together, whereas in the coactive (chart construction) task, group members rotated through subtasks so that at any one time they were working individually, but all group members worked on every part. The results showed that groups with high-ability leaders outperformed groups with low-ability leaders only in the coactive task. In the interactive (first) task, ability of the leader did not affect group performance. The investigators speculated that the leader had more control over group functioning in the coactive task than in the interactive task and so could make a substantial contribution to the group product. In the interactive task, one of the leader's responsibilities was to promote contributions by *all* group members, which would tend to deemphasize the leader's contribution. This conclusion is weakened by the fact that the rotation through subtasks confounds the coactive task. Rotation itself, or the fact that the leader participated himself in each phase, provides an alternative explanation to the finding of leader influence.

A third study, conducted at West Point by Adams, Prince, Yoder, and Rice (1981), also found a complex relationship between leader ability and group performance. In this study, cadets each led three-person groups on two tasks: a scale drawing of a building and writing a proposal for junior officers to maintain high standards and increase reenlistment rates. All groups were mixed-sex. Leader ability was defined by Scholastic Aptitude Test or American College of Testing scores. Only for the drawing task were any results significant. When the group leader was a male, leader ability was positively related to group performance when the group held traditional attitudes toward women,

but was *negatively* related to group performance when the group held liberal views. When the group leader was female, leader ability did not relate to group performance. Adams et al. did not attempt to explain the negative correlation, which is counterintuitive and puzzling.

Fiedler and Leister (1977) constructed a model that determined when leader intelligence should and should not be correlated with task performance. Their model included several mediating factors, called "screens," that should permit the correlation of intelligence and performance if they are not blocked by situational determinants. They examined this model by defining the various screens in terms of observable behavior, sorting out their population into medians on the screen, and then computing correlations between intelligence and performance ratings for each half of the sample separately. The subject sample was a group of staff sergeants rated by their superiors; intelligence was based on the Army entrance test battery. The most important screen, in producing different correlations for the two halves of the sample, was stress with a superior officer: in low stress, performance was strongly positively correlated with intelligence, but in high stress, zero or even negative correlations were found. Additionally, experience served as a mediator, with intelligence being more useful with more experienced leaders, with good leader-superior relations, and with good leader-group relations. Some indication that motivation, experience, and leader-group relations interact was also found: intelligence was correlated 0.58 or higher with performance for highly motivated, experienced leaders with good leader-group relations and also highly motivated, inexperienced leaders with poor leader-group relations, whereas strong negative correlations were found for highly motivated, inexperienced leaders and less motivated, inexperienced leaders, both with good leader-group relations. These findings are based on samples too small to have firm reliability (*n* for each group ranged from 9 to 23), but do suggest that the screens may operate in a nonstraightforward manner. Fiedler and Leister emphasize that stressors in the work situation work against intelligence helping performance; some of the negative correlations suggest that very stressful situations might be conducive to leader sabotage of superior officers' directives. Unfortunately, though, these authors do not present a model of *why* stress affects the intelligence/performance relationship; such a model would help guide the research needed to overcome the problems of stress.

Fiedler et al. (1979) continued the research on the relationship of intelligence, task performance, and stress in a series of four studies. Each study employed a different population of military leaders, ranging from infantry squad leaders to first sergeants, to Coast Guard staff, to

company commanders and batallion staff officers. In each study, intelligence, based on the GCT, was divided into medians or thirds, and the *performance vs. intelligence correlations were computed separately.* Each individual study has minor technical problems, but overall, the studies impressively demonstrate the main point that in situations of low superior-subordinate stress, the subordinate can effectively use intelligence to achieve good performance, but in situations of high stress, this relationship does not obtain. As before, experience was a mediating influence; the longer a person has served in a leadership capacity, the less were the negative effects of stress.

PERSONALITY OF THE LEADER

In addition to their work on leader ability, O'Brien and colleagues also related personality characteristics of the leader to group performance. O'Brien and Kabanoff (1981) and O'Brien and Harary (1977) tested the hypothesis that the discrepancy between the leader's need for control and participation and the opportunities for satisfying those needs would be negatively related to group productivity. In these studies, positions in the group were systematically manipulated so as to create congruence between need and opportunities (high need for control, high-control position; low need for control, low-control position) or discrepancy between need and opportunities (high or low need for control matched with low- or high-control positions). The tasks included building molecular models, writing stories from TAT pictures and abstract geometric shapes, discussions of general topics (e.g., capital punishment), writing Japanese Haiku poems, and interpreting Freudian dreams. Only for one task, writing stories from TAT pictures, was the relationship between discrepancy (of need for control and opportunities to apply control) and group performance significant. The greater the discrepancy between the leader's need for control and opportunities to apply control, the lower was group performance. Since most comparisons of leader discrepancy and group performance were not significant, there seems to be little support for O'Brien et al.'s discrepancy theory of group productivity.

Hewett, O'Brien, and Hornik (1974) examined the relationship between leader personality and group performance. In this study, in which the group task was to build models of molecules, the appointed leader was given instructions for the work organization to be used by the group (interaction or coaction). The task-orientation vs. the person-orientation of the leader was not related to group productivity,

nor did leader orientation interact with any other factor in the study (degree of interactivity, compatibility of the group).

In an independent examination of the personality of the leader, Gottheil and Vielhaber (1966) attempted to show how the interaction of leader and unit attributes related to the performance of military squads. Their sample was atypical: West Point cadets at a summer training camp during the annual Games Day. On this day, the different companies (assembled only for that summer) put their best squads in armor, artillery, signal, infantry, and engineering up against each other in performance contests. For each such squad, a leader is elected from the company. The authors attempted to ascertain how the interaction of characteristics of the leader and the rest of the squad affected squad performance. There were no differences in performance based on leader or member individual performance, aptitude for service rating, squad stability, or barracks type, either over event or for each event individually. This could be due to there being little room for improvement, given the personnel selected for the study. Differences between leaders and squad members were found for individual performance rating, aptitude for service rating, self-esteem, and degree of task motivation (leaders had more of each), and manifest anxiety (leaders had less). These differences, however, did not affect squad performance. It was found that the more the leader of a squad distanced himself from his erstwhile teammates (recalling that leadership was a Games Day election), the better the team performed. When squads had leaders with high esteem, presumably better able to distance themselves, the squad performed better when they took a task orientation and were critical of each other, whereas when the squad had a low-esteem leader, performance was better among squads rated as friendly. The authors interpret this finding in terms of cohesiveness: squads with cohesion can be more critical than squads lacking cohesion. This interpretation is difficult to understand given their data. What they conclude, although its immediate applicability to unit performance is unclear, is that leader and squad esteem are important factors in effective squad performance.

In summary, these studies show little evidence that the personality of the leader affects group performance. However, it must be kept in mind that the leaders in these groups rarely had "real" authority or power. The leader's primary responsibility in the laboratory experiments was to communicate instructions about the task, rather than to control the functioning of the group; in the military study, leadership was transitory. In natural settings, in which the leader has recognized and enduring authority, one might expect a stronger effect of personality on group functioning, and possibly on group performance as well.

Obviously, more research is needed before conclusions can be drawn about leader personality effects.

LEADER BEHAVIOR AND LEADERSHIP STYLES

Rather than review the voluminous literature on leadership styles and behavior, we will describe the conclusions drawn from the literature by Stogdill (1974), who reviewed hundreds of leadership studies. This description will be followed by a discussion of several studies, published after Stogdill's review, which used military men as subjects.

Stogdill reviewed a great number of studies that compared democratic vs. autocratic leadership, permissive vs. high-control leadership, follower-oriented vs. task-oriented leadership, high vs. low social distance leadership, and participative vs. directive patterns of leadership. Of all these comparisons, only social distance is consistently related to group performance: the greater the distance between leaders and followers, the higher the performance. Although conclusive statements about the relationship between leadership style and group performance cannot be directly assessed from a tabulation of results such as Stogdill's, it should be noted that many of the studies reviewed yielded statistically significant results. A closer scrutiny of the research data, employing meta-analytic statistical techniques and taking into account interactions among variables, might be a fruitful research project.

Klemp et al. (1977) obtained ratings of superior or average leadership performance for a sample of 82 Naval commissioned and noncommissioned officers based in San Diego (no rating of below average was possible). These officers were independently interviewed to identify critical incidents in their leadership experience in which they both succeeded and failed. These incidents were coded for the presence or absence of 27 separate leadership competencies, which in turn were factor analyzed, yielding five leadership factors:

- Orientation toward task achievement
- Skillful use of influence
- Use of management control techniques
- Advising and counseling
- Use of coercion

All factors but the last successfully discriminated between superior and average officers. Further analyses showed no major effects of the officer's service rank, whether he was commissioned, years of experience, and other variables. A cross-validation sample of men based in Norfolk showed that the factor structure and the discriminant function predicting leadership ratings could be replicated. It is possible that

leadership training could instill some of the qualities associated with effectiveness.

Yukl and van Fleet (1982) performed a multi-method, cross-situational analysis of military leader effectiveness. Subjects were either military cadets in a university program or Air Force officers. The method was either content analysis of critical leadership incidents elicited from subjects or correlational analysis of a leadership quality questionnaire. For each type of method, one noncombat- and one combat-oriented scenario were provided for analysis.[1] For both types of measurement and both combat and noncombat situations, four behaviors emerged as important for group performance:

- Emphasis on performance
- Inspiration
- Role clarification
- Use of criticism-discipline

There is a rough correspondence of these factors with those of Klemp et al. Emphasis on performance corresponds to task orientation, use of criticism-discipline corresponds to skillful use of influence, and role clarification corresponds to management control. There were some differences in combat vs. noncombat situations, with combat leadership showing more emergent problem solving. This could be due to the leadership experience of the officers or to the nature of situations that arise in combat, but not to personality difference because Yukl and van Fleet specifically excluded personality differences in leadership as variables in their study. The emergence of inspiration as a factor might arise from the method employed, which emphasized examples of outstanding leadership, rather than opportunities to be a good or bad leader.

The most prominent theory that takes into account interactions among variables, including characteristics of the task and environment, is that of Fiedler (1964, 1967, 1978). Fiedler's theory postulates that the type of leadership required for high group performance depends on the favorableness of the group-task situation for the leader, where favorableness refers to the ease with which the leader can influence group members.

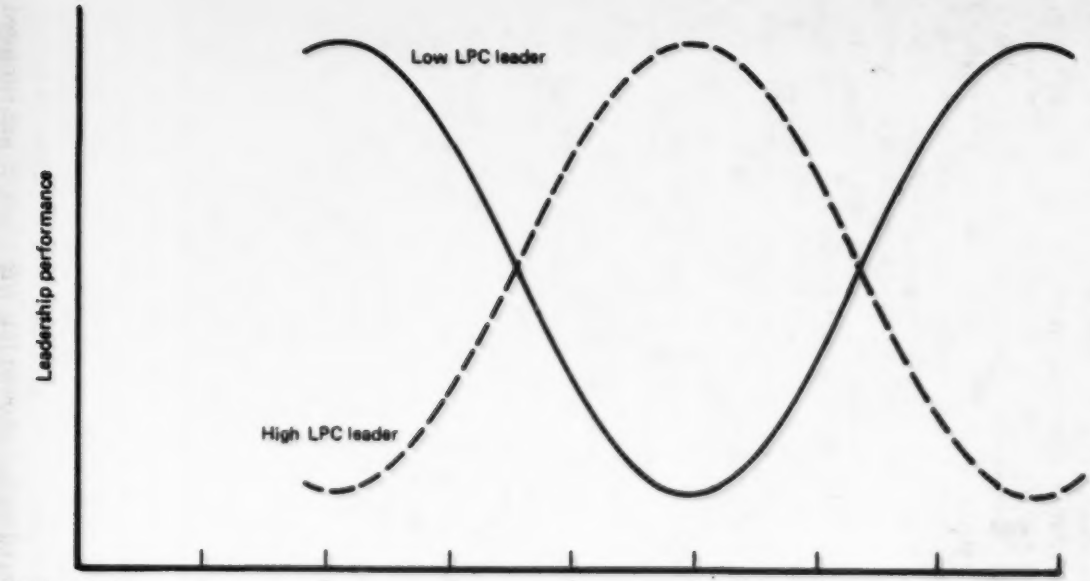
Leadership style is operationally defined in Fiedler's research as the "LPC" (Least-Preferred Coworker) score, which purports to measure an internally consistent, temporally stable personality trait. High-LPC

[1]The students were given the noncombat scenarios, while the officers, who had battle experience, were given the combat scenarios, so any difference on this dimension can be due to subject population, topic, or an interaction of the two.

leaders attend to the interpersonal problems in their group, whereas low-LPC leaders focus on the task to the neglect of group members. The relationship between effective leadership style and favorableness is complicated, as illustrated in Fig. 1, taken from Fiedler (1978). For extremely favorable or unfavorable conditions for the leader, directive and controlling leadership is expected to be most effective. When conditions are moderately favorable or unfavorable, permissive nondirective leadership is expected to be most effective. Shaw and Blum (1966, pp. 238–239) give a cogent and clear description of Fiedler's conditions of favorableness: the favorableness of the group-task situation is determined by three dimensions—the affective relation between the leader and his members, the degree to which the task is structured, and the power inherent in the leadership position. Although it is recognized that the interaction of these dimensions is complicated, Fiedler suggests that the leader's relation with his members is the most important structure of the task, and inherent power of the leadership position is least important for the favorableness continuum. A very favorable set of conditions is: leader-member relations are good, the task is highly unstructured, and the leadership has a high degree of inherent power. An unfavorable set of conditions has poor leader-member relations, an unstructured task, and a weak leadership position.

The prototypical military situation is a highly structured task with good leader-member relations and a strong leader position, and is most conducive to a task-oriented (low-LPC) leader. This finding corresponds to the emergence of the task-orientation factor as the strongest characteristic of good leadership in both Klemp et al. and Yukl and van Fleet, as discussed above.

Bons and Fiedler (1976) illustrate an application of Fiedler's model. One hundred fifteen infantry squad leaders were examined over a nine-month period. At the beginning of the period, their motivation (task vs. relationship) was assessed, along with the situational favorability of their environment, so that they could be placed into one of the eight categories indicated in Fig. 1. After the nine-month period, changes in their working conditions, including changes in assigned task, unit the leader commanded, and superior officers, were recorded and the situational favorableness was again assessed. It was shown that job changes brought about different changes in the person-related behaviors of leaders depending on whether they were task- or relationship-motivated and on whether the situational favorableness moved them to arenas more or less favorable to their particular leadership style. Experienced leaders showed this effect less; it was surmised that they are more used to frequent change. This study did not



Leader-member relations	1	2	3	4	5	6	7	8
	Good	Good	Good	Good	Poor	Poor	Poor	Poor
Task structure	High	High	Low	Low	High	High	Low	Low
Leadership power	High	Low	High	Low	High	Low	High	Low

SOURCE: Fiedler (1978).

Fig. 1—Fiedler's contingency model

directly address the effectiveness of the leaders, and also had acknowledged problems in its choice of subjects to sample, dropout rate from the study, and reliability of some measures. Nevertheless, it is, all of the problems included, a typical example of a mass of research based on the contingency model, nearly all of which has supported the model.

Although the support for Fiedler's theory seems impressive, Shaw and Blum (1966) point out that many of the studies supporting the theory derive leadership style from personality measures, rather than from observations of leadership behavior; when actual behavior is taken into account, results are mixed (see also Hare, 1976).

SUMMARY

No general statement can be made about how a leader's general ability, personality, and leadership characteristics affect group performance. In part, this lacuna exists because of the methodological problems inherent in the research, as discussed above. But we also cannot ascribe any direct effects on performance without knowing about many other aspects of the group's task, including the requirements of the task, the degree of structure in the environment, the cohesiveness of the group, the personality of the leader and group members, and the interpersonal compatibility of group members. Finally, we do not have much evidence that good leadership within a unit is a major contributor to good unit performance. This finding has the depressing implication that, for any new military unit task contemplated, a separate analysis of the task may be necessary to predict how leadership qualities affect unit performance.

V. GROUP STRUCTURE

From the review of individual characteristics, it was learned that at least for some tasks the presence of higher general or specific ability unit members leads to higher unit performance. But those results, addressing the effects of the ability of single individuals, do not tell us how to assemble groups from the available manpower, especially when individuals vary in ability. For example, when all group members contribute to a task, it is unclear whether a homogeneous group with moderate proficiency or a heterogeneous group with a wide range of proficiency will produce superior performance. To address such questions, we now move from consideration of individuals to the unit as a whole. The nature of groups, considered as a unit, has been of fundamental concern in social psychology (Carron, 1980), and thus we have had to be selective in order to limit the length of this review. Our major breakdown will be into studies reviewing (1) characteristics of group structure, or composition of the group, that affect performance, and (2) characteristics of the group process, or the interrelationships of members. We begin with group structure. We will not review here the effects of environmental influence on group performance except as those influences impinge on group structure or process; it should be kept in mind, however, that such influences can be of major importance (see, for example, Marks and Mirvis, 1981).

The literature predicting group performance from the structure of the group has focused on such group characteristics as size and turnover rate as well as general ability, proficiency on the task, personality characteristics, and interpersonal compatibility. Although not every study is reviewed here, the studies included represent a wide range of tasks and rules for composing groups (homogeneous groups at different levels—high, medium, and low—and heterogeneous groups). Particular attention is paid to tasks requiring motor manipulation and physical coordination among group members.

SIZE AND TURBULENCE

A Naval study has examined the subjective size of a unit as it relates to unit performance, while an Army Research Institute project has examined the effect of turnover rate on performance. Dean et al. (1979) examined the influence of size of the group on its performance in a study of Navy crews. Their basic hypothesis was that when unit

members feel that manning levels are sufficiently high, performance will be better than when the subjective impression is one of understaffing. In this large-scale study, shipboard crews were given a battery of items (the "Shipboard Habitability and Climate Questionnaire") at the beginning of a six- to eight-month tour of duty. The questionnaire assessed many different aspects of shipboard life, including quality of living conditions, perception of manning levels, and perceived work effort required. In addition, illness records, age, pay grade, and a number of other measures were obtained for each subject, as well as manning information for different work groups of each ship surveyed. The productivity measures were subjective ratings made by department heads, assessing separately the dimensions of competence, maintenance, readiness, stress, efficiency, cooperativeness, safety, and, for petty officers, leadership. The sum of ratings was the major dependent variable of the study. Data were analyzed on the individual sailor and shipboard department level, using regression techniques. Dean et al. report that the actual manning levels did not predict productivity well, but the manner in which manpower is perceived to be utilized was important. This latter was a composite of items from the major questionnaire assessing such items as perception of matching abilities to jobs, extent of friction within crews, extent of work assistance available, and pride of workmanship. It was concluded that manpower utilization perceptions of efficient use of human resources should be closely monitored, as these perceptions influence group performance.

An element of group structure that has caused considerable concern in the military is the replacement of personnel. Folk wisdom holds that increased turbulence causes a decrement in unit performance, as time is required for individuals to learn each others' habits and to function effectively as a team. Carron (1980) reviews the sports literature in this regard, and shows that there is a relationship between turnover in major sports teams and poor performance, but the causal direction for that finding is far from established.

The question has been directly addressed in a study of tank crew stability as it relates to tank gunnery performance (Eaton, 1978b); Eaton and Neff, 1978). In the first study (Eaton, 1978b), questionnaires were given to 248 tank crews to determine how long the crew had served as a unit and how long each member had served in his particular role. The 198 usable questionnaires were used to predict performance on the Table VIII tank gunnery exercise. Turbulence was defined in terms of (1) length of time each crew member had been in his particular role, (2) length of time the crew had served together as a unit, and (3) length of time the unit had been together with their particular tank. For group measures, the tank commander provided

responses; crew members answered items about themselves as individuals. The various turbulence and Table VIII score measures were intercorrelated with each other; in general, any significant relationships were weak, in the range of 0.19 to 0.28 in absolute value. The statistically significant relationships were exclusively individual ones: the experience of the gunner was positively related to the number of targets hit; and the experience of the tank commander was related to main gun opening time. No effect for team turbulence was found. It should be noted that the length of experience for all team members except the tank commander was fairly short, which may have attenuated the length of time the crew could serve together as a team, and thus weakened any correlations with performance. As the next study will show, however, this is unlikely.

Eaton and Neff (1978) extended the previous study to an experimental analysis of turbulence. As a control condition, they employed the intact tank crews of the previous study. This first condition was compared with three experimental conditions especially created for the study. In condition 2, crews were mixed so that each member was in his correct role, but the four members (tank commander, gunner, loader, and driver) came from different units. In this condition, experience in role was controlled, but turbulence varied, as the unit was artificially created and therefore brand new. In condition 3, experience as well as turbulence was altered. Here, gunners served as tank commanders, and loaders became gunners (drivers remained drivers, and second loaders served as loaders). Finally, in the fourth condition, the tank commander and driver remained together, and nonarmor personnel, after a three-day training session, served as gunners and loaders. Conditions 1 and 2 did about the same, and both did significantly better than condition 3. This demonstrates that individual experience, and not the intactness of the team, is an important factor in tank crew performance. Finally, condition 4 did surprisingly well, surpassing condition 3 in performance, and scoring only slightly below conditions 1 and 2. This indicates that the roles of tank commander and driver are central to tank crew performance, with gunnery and loading duties that can be rapidly learned.

The series of studies by Eaton and his co-workers indicate that turbulence might not be as important a factor in tank crew performance as was believed. The experimental study showed that intact crews do not outperform *ad hoc* assembled crews, and the field study showed that the length of time an intact crew was together was only a mediocre predictor of performance. However, because intact crews in both studies had generally been intact only for a short time, the possibility that extended time together could influence crew performance cannot

be ruled out. A true test of this hypothesis would call for a major extension of the Eaton and Neff study in which crews were randomly assigned to either lengthy (1 year or more) duty together or shorter periods together with crew shifts. This would be a complicated and expensive study, and whether the benefit would outweigh the cost is not clear.

DISTRIBUTION OF GENERAL ABILITY[1]

Although it seems intuitively obvious that a group's performance depends on the distribution of abilities of its members, there is more theoretical discussion than empirical research on this topic. A theoretical formulation by Steiner (1972) is useful for showing when heterogeneity of abilities is likely to be advantageous for group productivity and performance. Steiner developed a catalogue of five types of tasks, and discusses for each the effects of heterogeneity of member ability on group performance. This formulation posits that group composition will interact with the requirements of the group task; for some tasks, heterogeneity of abilities is beneficial for group performance, whereas for other tasks, heterogeneity is detrimental. For still other tasks, the range of abilities in the group is irrelevant to group productivity.

Steiner's catalogue includes the following types of tasks: disjunctive, conjunctive, additive, discretionary, and divisible. In *disjunctive* tasks, group performance is determined by the ability of the group's most competent member. The task most often cited is rope-pulling, where each team nominates one person to pull the rope. If two teams have equal means, the more heterogeneous team will win. In *conjunctive* tasks, the group's performance depends on the ability of the least competent member. A team of mountain climbers, for example, can proceed no faster than its slowest member. When two teams have equal mean ability, the more homogeneous will move faster. In *additive* tasks, group performance is an additive combination of all group members' abilities, as in team rope-pulling in which all members of the team pull the rope. Since group performance is expected to be a function of its total pulling power, the heterogeneity of abilities is irrelevant. In *discretionary* tasks, members of a group combine their efforts in any way they choose. When the task is to correctly estimate the distance of an object, for example, the group may elect to pool

[1]Note that in this subsection we are primarily addressing the distribution of member ability, which is a feature of group structure. To some extent, we will also consider mean overall ability, in a partial overlap with the discussion of the contribution of individual member ability, above.

members' judgments or to use the judgment of the most competent or experienced member. For two groups with equal means but different distributions of abilities, the more heterogeneous group has the potential for making a more accurate judgment than the homogeneous group, but this potential will be realized only if members' judgments are weighted by their ability. In *divisible* tasks, different group members perform different subtasks. The group's performance depends on the distribution of specialized skills within the group. If the task is to build a bridge, for example, a group with specialists in design, engineering, and construction will produce a better product than a more homogeneous group in which all members possess some skill in each area.

Although the above conclusions are sensible in the abstract, many tasks fall into multiple categories, or are modified by the setting in which they are performed. For example, a wartime setting imposes additional constraints on task performance. Although bridge-building teams with specialists in different subtasks may in theory be expected to be superior to teams in which all members have some skill in each area, the latter groups may be more successful in the event that one or more members become incapacitated. Such considerations have rarely been taken into account, either in theoretical discussion or in designing empirical research.

The few studies that have contrasted heterogeneous and homogeneous groups on member ability have produced inconsistent findings. This is not surprising, however, because the tasks used in these studies have different requirements according to Steiner's (1972) scheme. Not only are the tasks different, but the measures of ability also differ across studies, making comparisons difficult.

For example, an early investigation of ability composition conducted by Shaw (1960; see also Shaw, 1981) correlated the average deviation among group members' scores on the Scholastic Aptitude Test and group performance on intellectual problem-solving tasks. None of the correlations between heterogeneity of ability and group performance was statistically significant. This result is not conclusive, however, since it is not known how group members combined their resources, nor is it clear whether the ability measure (SAT scores) was a good proxy for the skills needed to solve the problems.

Another series of studies investigated group heterogeneity in creative ability. Triandis and colleagues (Triandis, 1959a, b, c, 1960a, b; Triandis, Hall, and Ewen, 1965) formed groups on the basis of creativity and attitudes (for example, conservatism-liberalism). The study with the most complete design formed all possible combinations of dyads on creativity and attitudes (homogeneous-liberal, homogeneous-conservative, and heterogeneous on attitudes crossed with

homogeneous-creative, homogeneous-uncreative, and heterogeneous on creativity). The tasks given to groups were intellectual problems (e.g., how can a person with no particular talent achieve fame, or how can a church in a poor neighborhood obtain funds to complete its building). Independent judges rated the number and quality of the group solutions. Although the results were complex, the major finding was the following: dyads that were heterogeneous in attitudes and homogeneous in creative ability produced more high-quality solutions than (1) dyads that were heterogeneous in both attitudes and ability and (2) dyads that were homogeneous in both attitudes and ability. This result suggests that one characteristic (here, attitudes) can moderate the effects of another (ability).

Two other studies compared group compositions on the basis of ability and attitudinal similarity. Ability was varied across groups, but remained homogeneous within groups, while attitudinal homogeneity was varied. The first study, comparing homogeneous groups, is often cited as a good example of a divisible task, in contrast to the purely intellectual problem-solving tasks typically studied. Of all the tasks described in this section, it is probably the most similar to small-group tasks in military settings. Terborg, Castore, and DeNinno (1976) used field projects in land surveying. Undergraduate students in a course on land surveying worked in three-person groups on three surveying projects. Each project contained three subtasks: operating the plumb line, working the transit, and writing down the results. Students rotated across the positions on different projects. As Terborg et al. noted, the task was not only divisible, but it was also additive (a group's performance was the sum of the three subtasks) and partially disjunctive (the performance of the group was heavily influenced by the person operating the transit). Group composition was determined on the basis of general ability (a combination of SAT scores and grade point average) and attitudes toward general topics such as state income tax, legal drinking age, and athletics (Survey of Attitudes Questionnaire, Byrne, 1971). All groups were homogeneous on ability, half of the groups had above-average ability, half had below-average ability. Within each ability type, half of the groups had similar attitudes across group members, and half were composed of group members with different attitudes. Ability and attitude similarity were expected to be positively related to group performance. Not surprisingly, high-ability groups outperformed low-ability groups. Attitude similarity had no effect on performance. The comparison between homogeneous high-ability and low-ability groups was not very informative, however; more useful comparisons would be (1) homogeneous vs. heterogeneous groups controlling for

mean ability, and (2) comparisons among different ranges of ability within the high-ability and low-ability categories.

In another study comparing homogeneous groups at different levels, Sorenson (1973) formed groups that were homogeneous-high or homogeneous-low in creativity. Half of the groups were also high in cognitive social differentiation; half were low on this characteristic. Sorensen operationalized cognitive social differentiation by asking participants to rate other people on abstract dimensions, such as creativity, and classified subjects as high- or low-differentiators according to the number of points in the scale they used. Two types of tasks were used: creative writing and intellectual problem-solving; each task was scored on the basis of quality and originality of the solution. Although one might expect that groups that were high on both creativity and differentiation would outperform the other groups, this was not the case. Group performance was highest when group members were high on only *one* trait. Groups that were high on either creativity or differentiation outperformed both groups that were high on both traits and groups that were low on both traits. This result was consistent across tasks and performance measures. Sorenson's examination of group process suggests that the groups that were high on both dimensions were so critical of each others' ideas that they had difficulty arriving at final solutions to the tasks. This interesting result shows (1) that ability characteristics may interact in unexpected ways, and (2) that it is important to try to clarify such unexpected and complex results.

In summary, there is no general conclusion about the most optimal group composition on general ability. The most interesting results, found in several studies using different tasks, is that groups composed of all high-ability members do not necessarily perform better than groups composed of members with moderate ability or with a range of abilities. Since ability interacted with other factors, such as attitude and cognitive style, homogeneous high-ability groups performed best only if they were heterogeneous or low on other factors. This interaction indicates that ability is not the sole dimension affecting productivity, and that factors entering into the non-task-oriented aspects of the group (such as attitudes) can be important moderators of the effects of ability mix. The implications of this indication will be explored below.

TASK-RELATED ABILITIES AND PROFICIENCY IN THE TASK

We have discussed group composition on the basis of general ability variables, including general scholastic aptitude and creative ability.

Two difficulties in interpreting the results are (1) although groups were composed on a particular ability variable, the observed differences between group compositions may have been due to other ability variables not measured, and (2) general ability variables typically have low correlations with task performance even at the individual level. Less ambiguous interpretations would arise when groups are composed on the basis of task-related abilities or proficiency. A number of studies have compared group compositions using an ability measure that is likely to be strongly related to performance: proficiency in the *same* task to be performed in the group setting. We now review the results of several of these studies.

Goldman (1965) individually administered college students the Wonderlic Intelligence Test and then formed the following dyads on the basis of the results: high-high, medium-medium, low-low, high-medium, medium-low, and high-low. These dyads then retook the test as a team to produce one set of answers. Although Goldman was primarily interested in comparing individual and group performance, and did not design the study to compare heterogeneous and homogeneous groups, such a comparison is possible. As Shaw (1971) pointed out, pooling all heterogeneous groups into one category and pooling all homogeneous groups into another produces two categories with approximately equal means at the outset. The heterogeneous category performed significantly better on the group task than the homogeneous category, suggesting that heterogeneous groups are more effective than homogeneous groups.

A similar series of studies was conducted by Laughlin and colleagues (Laughlin and Branch, 1972; Laughlin, Branch, and Johnson, 1969). In these studies, triads (the 1969 study) and tetrads (the 1972 study) completed the Terman Concept Mastery Test first individually and then in groups. Laughlin and colleagues formed all possible combinations of ability: group compositions ranged from all high-ability to all low-ability. As in the Goldman study, pooling all homogeneous groups and pooling all heterogeneous groups produced two categories with nearly equal means on the pretest. On the group task, the heterogeneous category outperformed the homogeneous category. A particularly dramatic comparison is that between homogeneous medium-ability triads and heterogeneous triads with a high, medium, and low. Although the means of the two groups at the outset were nearly identical, the homogeneous groups achieved a mean of 49.94 on the group task, whereas the heterogeneous groups achieved a mean of 63.75 (the maximum possible was 115). A similar result occurred in Laughlin and Branch's (1972) study of tetrads although the effect was not as pronounced.

The advantage of heterogeneous over homogeneous grouping in the Goldman (1965) and Laughlin et al. (1969, 1972) studies makes sense since the intellectual task can be seen as a combination of Steiner's (1972) disjunctive task, in which group performance is a function of the proficiency of the most competent group member, and a discretionary task, in which group members may pool individual resources any way they want. In both cases, heterogeneity of abilities is more advantageous than homogeneity.

The tasks in the above studies involved pooling of resources but did not require true coordination of efforts. Furthermore, they were purely intellectual tasks. The final two studies reviewed here are important because the task involved motor abilities and required group members to physically coordinate their efforts.

Gill (1979, p. 115) used a motor maze task (described in the discussion of individual characteristics) in which two members of a dyad operated different handles that tilted the maze board. Individual proficiency was the average time of ten practice trials and group proficiency was measured during ten group trials. Heterogeneous dyads were formed on the basis of individual proficiency so that the difference between partners' proficiencies was at least seven seconds; homogeneous dyads each had a range of four seconds or less (individual proficiency ranged from 19.7 to 57.7 seconds). The mean initial proficiency of the heterogeneous groups was the same as that of the homogeneous groups. To determine the effect of group composition on group performance, Gill performed multiple regression analyses of group performance with mean group proficiency and the intradyadic difference in performance as the predictors. Not surprisingly, average proficiency was positively related to group performance: the higher the group mean, the higher was the group's performance. Interestingly, however, the difference between members' proficiency was negatively related to performance even when group mean proficiency was taken into account. In other words, for groups with the same mean proficiency level at the outset, groups with a wide discrepancy between individual members' proficiencies did *worse* on the group task than did groups with a narrow discrepancy between individuals' proficiencies. As Gill described the results, the proficient partner could not compensate for the other partner's poor performance. This finding has a parallel with team tasks in the military. In tank crew performance, for example, a highly proficient gunner cannot compensate much for a poor driver. On the basis of Gill's results, a crew whose members all have medium

proficiency would be predicted to be more effective than a crew in which some members have high proficiency and others have low proficiency.

Jones (1974), in the study discussed earlier on ability in major league sports, conducted a different type of analysis on his data to examine the issue of arranging teams so as to maximize performance *across* teams. That is, if we wish to maximize the performance of a collective of teams, does it matter how we assign members to teams? For example, is it better to have two teams, one made up of the best players at each position and one made up of the worst players at each position, or is it better to mix ability levels, if our objective is to obtain the best possible aggregate score over the two teams? Jones examined this question in two manners, one asking how professional teams actually distributed ability, and the other asking whether it made a difference. The answer to the former question depended on the sport. In tennis, baseball, and football, good players tended to team with good players. In basketball, on the other hand, good players tended to be isolated on different teams. This is most likely a function of the individual sports, including how they are attractive to audiences. The second finding was that the summed effectiveness over teams did not depend on how the constituents were assembled. Jones based this conclusion on examination of a term in the prediction equation for performance that measured the interactiveness of the individual members' abilities. This measure has the effect of assessing any performance differences due to differential ability other than the sum of player abilities. For all four sports, this term did not contribute to the predictiveness of the model. This result implies that, given a fixed group of potential crew members, you cannot improve total performance by using an ability measure to assign members to subteams.

In summary, the studies assembling groups on the basis of member proficiency on the task have found heterogeneous groups to be superior to homogeneous groups on intellectual tasks where group members could pool members' resources in any way they chose. For tasks requiring true physical coordination among group members, however, heterogeneity seems to be detrimental to group performance or to have no effect. Because so few studies have been conducted using tasks requiring coordination among members but also permitting some flexibility in how group members pool their skills, the relationship between heterogeneity and group performance for this important class of tasks remains to be investigated.

HOMOGENEITY OF PERSONALITY AND COGNITIVE STYLES

A number of studies have compared heterogeneous and homogeneous groups on personality characteristics and cognitive style. The personality characteristics examined include general personality profiles, supervisory ability and decisionmaking approach, concrete vs. abstract problem-solving styles, reflective vs. active problem-solving styles, and interpersonal effectiveness. As we shall see, the evidence on the superiority of homogeneity vs. heterogeneity is mixed, depending on the characteristics measured and task performed. Here, we shall examine personality and cognitive characteristics that are related to the task, whereas in the following subsection, we will examine personality, cognitive, and social characteristics of the group that are related to its social composition.

Hoffman and colleagues (Hoffman, 1959; Hoffman and Smith, 1960; Hoffman and Maier, 1961) have compared personality profiles of group members to group performance. For example, Hoffman and Maier (1961) used the Guilford Temperament Survey to measure ten personality variables, and formed groups whose members had similar profiles (homogeneous groups) or whose members had different profiles (heterogeneous groups). Groups worked on four discussion problems (e.g., develop a method for permitting five men to cross a heavily mined road, or decide how to allocate funds from a limited source to needy students). Independent judges rated the quality of the groups' solutions to these problems. On three of the four problems, heterogeneous groups scored higher than homogeneous groups. On the fourth problem, the two kinds of groups performed equally well. Hoffman and Maier hypothesized that heterogeneous groups were superior because they represented diverse problem-solving perspectives. Although their interpretation is reasonable, Steiner (1972) points out that heterogeneity of personality is not necessarily correlated with heterogeneity of viewpoints. Another interpretation of this result is that groups with varying member personality profiles are more compatible than groups in which all members exhibit the same personality characteristics. This issue will be discussed below in the subsection entitled "Interpersonal Compatibility."

The studies comparing group composition based on single personality characteristics generally agree with Hoffman et al.'s conclusions that group heterogeneity is superior to group homogeneity, although the explanations for the results vary from study to study. Lampkin (1972), for example, compared five group compositions using need for dominance (need to assert influence over others), one of the variables

in the personality profiles used in the Hoffman et al. studies. The five group compositions were homogeneous-high, homogeneous-medium, homogeneous-low, two high and one low, and one high and two low. The task given to triads was a consensus decisionmaking problem. Group members were shown three visual patterns on a television screen, in which two of the three patterns were identical and the third differed in minor detail, and were asked to reach a consensus decision about the matching patterns. Since participants in the study had been trained previously to perform the task individually at a high level of accuracy, the accuracy of decisions was constant across groups, ranging from 75 to 80 percent correct. The measure of group performance was, therefore, the time taken to reach a decision. In all sessions of the study, heterogeneous groups reached consensus significantly faster than homogeneous groups. Not only was this true on the average, but also the two categories of group composition showed nonoverlapping distributions: the slowest heterogeneous group composition was faster than the fastest homogeneous group composition.

Lampkin (1972) offered the following explanation for the relative inefficiency of homogeneous groups: homogeneous high-dominant groups spent much of their time trying to change each other's opinions, and homogeneous low-dominant groups spent time ascertaining each other's opinions without trying to achieve consensus. In contrast, heterogeneous groups spent relatively little time communicating and reached consensus the fastest. In Lampkin's study, since accuracy was held constant, group efficiency in decisionmaking could only be interpreted as a desirable outcome. In real settings, however, there may be a tradeoff between the time taken to reach a group decision and the accuracy of the decision. The effect of group heterogeneity in its members' need for dominance on the optimization of speed and accuracy has not yet been investigated in settings where accuracy is free to vary. In decisions of timing (Rapoport et al., 1976), where the decision of *what* to do is less crucial than the decision of *when* to do it (such as firing a SAM), Lampkin's results may be applicable.

Two often-cited studies have compared groups with different distributions of supervisory ability and decisionmaking approach (Ghiselli and Lodahl, 1958; Lodahl and Porter, 1961). Supervisory ability and decisionmaking approach are two scales of Ghiselli's Self-Description Inventory (Ghiselli, 1954). The first scale differentiates between persons believed adequate for supervisory responsibilities and those believed inadequate, and the second scale differentiates people on such characteristics as self-reliance, general activity, and willingness to take action based on their assessment of the situation and their own abilities (see Porter and Ghiselli, 1957).

These studies are particularly noteworthy because they used nonintellectual tasks. In the Ghiselli-Lodahl study, groups operated a model railroad which had two trains going in opposite directions. There were two sets of electrical control panels with switches that controlled the trains and track. Group performance was based on the number of times *both* trains circled the track without wrecks or derailments. The Lodahl-Porter study examined intact groups of industrial workers (predominantly airplane maintenance crews) performing their usual jobs.

Both studies found that the distribution of supervisory ability and decisionmaking approach in a group influenced group productivity. However, the results of the studies were in opposite directions. In the Ghiselli-Lodahl study, neither the mean score nor the highest score in the group on supervisory ability or decisionmaking approach related to productivity. The difference between the two highest scores and the positive skewness of the scores in the group (for both scales), in contrast, were positively related to productivity (correlations ranged from 0.44 to 0.82). Thus,

> whether or not the group contains a person who stands high in approach to decision making, that is, tends to be self-sufficient in decision making, is not significantly related to group performance. However, if the group possesses such a person and his position in the group on this trait is relatively uncontested, then the group is likely to be superior in productivity.... When the group possesses an individual who is uncontested in this trait and the remaining members of the group are homogeneous with respect to it, then there is a very high degree of likelihood that the productivity of the group will be high. (Ghiselli and Lodahl, 1958, p. 64.)

This same conclusion applies to supervisory ability. As Steiner (1972, p. 120) summarized, "Heterogeneous groups were more successful than homogeneous ones, but it was only the difference between the top member and all the rest that really mattered."

The significant findings in the Lodahl-Porter study contradicted those of the above study. Lodahl and Porter correlated the productivity of airplane maintenance crews with the group mean, heterogeneity of scores within the group (the standard deviation), skewness, and the leadman's percentile position in the group. The leadman in a group was a mechanic with high seniority in the company who was judged by management to have considerable influence over other team members. As in the Ghiselli-Lodahl study, the group mean was unrelated to productivity. The similarity between the two studies ends there, however. Positive skewness and the leadman's percentile position in the group were both *negatively* related to productivity, as well as

the heterogeneity of supervisory ability within the group. The correlations for heterogeneity and leadman's percentile position were statistically significant, whereas skewness approached significance. Lodahl and Porter explained these findings by suggesting that heterogeneity of supervisory scores is associated with low cohesiveness, which is associated with low productivity (cohesiveness and productivity correlated 0.54). Furthermore, they suggested that the supervisory ability of the leadman was negatively related to his popularity, which in turn was related to productivity (the leadman's popularity correlated 0.64 with productivity). This interpretation is related to a major distinction between the two studies: ad hoc groups that functioned for a short time vs. intact groups that functioned together for months or years. The impact of heterogeneity of supervisory ability and decisionmaking approach (as well as other characteristics) may differ in the two settings. Certainly, it is doubtful that cohesiveness and popularity (possible mediating variables between supervisory ability and productivity) have major influences on the short-term functioning of ad hoc groups.

The third dispositional characteristic used to form groups is concrete vs. abstract problem-solving style. Tuckman (1964) defined four levels of this variable ranging from very concrete to very abstract, and then composed homogeneous groups at each level. Groups then played a stock market game which is "an emulation of the stock exchange where teams buy and sell stocks and bonds in order to accumulate more profit than their competitors" (Tuckman, 1964, pp. 478–479). Groups were awarded points and cash prizes according to their ranks on net accumulated gain. In keeping with the above findings on individual personality characteristics, there was no relationship between abstractness of group members and group performance; profit could be gained by using either concrete or abstract strategies.

Tuckman (1967) conducted a second study employing the psychological dimensions of dominance and abstractness/concreteness to examine heterogeneity. Two military-type tasks were used, one a structured object identification task and the other an unstructured hypothetical tactical exercise. Twelve three-man groups of Navy enlisted men volunteered as subjects, performing both tasks. Groups were homogeneous on none, one, or both of the two dimensions. As expected, in the unstructured task, groups in which abstract subjects were the majority outperformed groups with a majority of concrete members. The reverse did not hold true on the structured task, however, as there were no differences due to independent variables. The groups intermediate in homogeneity (mixed on dominance, but homogeneously abstract) did best on the unstructured task, but worst on the structured task. In general, no evidence for superiority of homogeneous groups

was found. Tuckman conjectured that an intermediate level of heterogeneity suppresses the formation of a group structure which would aid in structured and hinder in unstructured tasks, but how homogeneous and heterogeneous groups can both form structures while intermediate ones cannot is not explained.

Lord and Rowzee (1979) did not compose groups in a particular way but instead correlated the heterogeneity (standard deviation) of groups on abstract vs. concrete problem-solving styles with group performance on four tasks. The tasks included cryptograms, pairing statements with implications of the statements, sorting a deck of playing cards into groups with different sums, and constructing sentences from words written on individual cards. Consistent with Tuckman's (1967) result, none of the linear correlations between heterogeneity and performance was significant (the highest correlation was 0.10); unfortunately, the authors did not test for the curvilinearity Tuckman did find.

The same study also examined the relationship between heterogeneity of reflective vs. active problem-solving styles and group performance on the tasks described above. Contrary to the results for abstract vs. concrete problem-solving styles, Lord and Rowzee obtained a significant negative correlation for the cryptogram task. The greater the standard deviation of this problem-solving style in the group, the poorer was group performance. The investigators attributed this result to communication difficulty in groups with a wide dispersion on problem-solving style. On a post-experimental questionnaire, participants reported on their group's difficulties in communication. Heterogeneity on the reflective/active dimension was positively related to communication difficulty. Heterogeneity on the abstract/concrete dimension, in contrast, was not related to communication difficulty, which the investigators used to help explain the different results for the two problem-solving styles.

Roberts, Meeker, and Aller (1972), in their examination of Naval officers who attributed causality to force, strategy, or both (see Sec. III on individual personality measures), found that heterogeneous groups had better performance in a decisionmaking game than did homogeneous groups. Apparently, having a variety of opinions about the structure of the problem allowed for more successful performance.

The final characteristic, interpersonal effectiveness, was studied by Bouchard (1972). Bouchard defined interpersonal effectiveness as the sum of the first five scales of the California Psychological Inventory: dominance, capacity for status, sociability, social presence, and self-acceptance. Homogeneous groups were formed that were high or low on this composite. Groups were asked to brainstorm on names for a new toothpaste, uses for an old tire, and uses for an extra opposable

thumb. The group's ideas were scored according to number and quality. On only one task and for only one outcome measure was there any effect for group compositions: high interpersonal-effectiveness groups produced ideas of higher quality for the opposable thumb problem than did low interpersonal-effectiveness groups. This significant result was explained on the basis of the well-developed social skills, verbal fluency, and outgoingness of persons scoring high on the interpersonal-effectiveness measure. The lack of significant results for the other tasks, however, raises doubts about the reliability of the significant result.

In summary, the studies relating personality composition to group performance demonstrate that the particular measures of personality, task requirements, and group composition each affect results in major ways. The studies that reported heterogeneous groups to be superior to homogeneous groups used general personality profiles to predict performance on discussion problems, need for dominance to predict group decisionmaking performance, and supervisory ability and decisionmaking approaches to predict group perceptual-motor task performance. The studies showing homogeneous groups to be superior to heterogeneous groups predicted group perceptual-motor task performance from supervisory ability and intellectual problem performance from reflective vs. active problem-solving style. The studies showing no difference in performance between heterogeneous and homogeneous groups examined abstract vs. concrete problem-solving styles for intellectual problems and games. Such a large range of outcomes of variables indicates that any novel personality-task combination should be investigated in its own right.

INTERPERSONAL COMPATIBILITY

Closely related to group composition on the basis of personality and dispositional characteristics is the compatibility of group members. In fact, many researchers and reviewers use the terms interchangeably. Here, we shall concentrate on those aspects of homogeneity and heterogeneity of group structure that have an impact on intragroup relations; in this sense, interpersonal compatibility serves as a bridge between group structure and group process.

A number of sports psychology studies have addressed social and personal homogeneity and heterogeneity of group members as they influence performance. A review article by Eitzen (1978) expresses the folk wisdom of the field when it states the viewpoint that more homogeneity in a group leads to positive bonds, which in turn lead to better

performance in interactive groups. Heterogeneity leads to cliques and separation of the group. Eitzen cites several studies showing that player turnover in major sports such as European soccer and American major league baseball are associated with less success, but he does not emphasize that the causal connection between these two phenomena is not at all clear.

One of the studies supporting this viewpoint most strongly is by Eitzen (1973), in which he assessed the social homogeneity and success of basketball teams from small high schools throughout the state of Kansas. In this study, data were obtained from coaches of small Kansas high schools (less than 700 enrollment). Of the 366 coaches who were queried, 288 responded. Respondents were asked to provide data on their starting five players, giving race, father's occupation, status in town (high vs. low), religion, and distance of residence from town (in town, out of town). Homogeneity was defined as four out of the five players the same for the status, residence, and religion questions, and below the median in absolute distance from the mean for the parental socioeconomic status question. Race was discarded because so few Kansans in small towns were not white. For each dimension of homogeneity, a chi-square analysis setting success (winning more games than losing) against the homogeneity measure was performed. In addition, coaches were asked to indicate the extent to which their starting five players belonged to cliques. This measure, also, was compared to the four homogeneity measures. It was found that for each of the four measures, heterogeneity was associated with increased incidence of cliques, and, moreover, the more dimensions on which the group was heterogeneous, the more likely were cliques.[2] Using only the raw measures of homogeneity, only homogeneity of family status predicted winners, but a summary measure of number of dimensions of homogeneity was monotonically related to the likelihood of being a winning team. A breakdown of teams into those with and without cliques showed that for teams without cliques, the more homogeneity, the more a team won, but for teams with cliques, homogeneity was not related to success. This study, then, indicates that the more a team is socially homogeneous and free of divisive factions, the better it can function.

Eitzen's results are not confirmed by other studies. Melnick and Chemers (1974) examined the degree of status homogeneity in 21 university intramural basketball teams, and found no correlation between these pretournament measures and performance in the basketball season. Foeldesi (1976), in a small-sample intense study of the

[2]This could be because of collinearity; the correlations among the various dimensions of homogeneity were not reported.

Hungarian national rowing team, found no evidence that homogeneity of socioeconomic status affected performance. Klein and Christiansen (1968) found the contrary result that heterogeneous teams, with respect to need for achievement, outperformed homogeneous teams in West German basketball teams. All these studies began with a belief in the homogeneity-leads-to-superior-performance hypothesis, so their results were "unanticipated." However, these essentially negative results might be attributable to a selection bias, such that only individuals predisposed to team commitment play in major sports competitions.

Altman and Haythorn (1967) examined the effects of social homogeneity on four personality dimensions: need for achievement, need for affiliation, dogmatism, and need for dominance in a small-sample complicated experiment using volunteer Navy recruits. Thirty-six subjects were paired into dyads who were both high, both low, or heterogeneous with respect to the four personality dimensions in a Greco-Latin square design.[3] Performance was measured on an individual vigilance task and on two group tasks requiring interaction. In addition, half of the dyads were examined in conditions of social isolation, where they lived together with no other social contact, whereas the other half lived on a Navy base. This isolation condition, extending earlier work that demonstrated that individuals in social isolation show strong performance decrements, was of major interest to the researchers. Results showed a slight decrement in performance on the individual task for the isolated dyads, but enhanced performance on the group tasks. Moreover, the anticipated enhancement of performance with homogeneity was not obtained; in general, heterogeneous dyads outperformed homogeneous ones of either level on the personality dimensions.[4] Thus, this study also argues against the homogeneity-leads-to-improved-performance hypothesis, and is free from the subject selection bias that affected the sports team studies. However, the limitation to groups of size n=2 restricts any generalizability to larger groups (Rapoport, 1971).

In addition to social homogeneity, which measures closeness between individual group members, status congruency, which measures the extent to which group members are ranked similarly on different dimensions, has been studied. To illustrate the concept, status

[3]Such a design permits one to use nine groups to examine the four dimensions independently. But it must be assumed that the various personality dimensions do not interact with each other; for example, one must assume that there is nothing about a group heterogeneous in both dogmatism and need for dominance that does not arise from considering each dimension separately. Such an assumption is at best questionable.

[4]Subjects were deliberately matched on age, size of hometown, education, religion, and family size, so heterogeneity may have been a way for subjects to maintain a sense of individuality.

congruency exists in a military organization if families of higher-ranked enlisted men are higher in social-economic status. It has been shown (e.g., Kahan and Poitou, 1973) that people will create status congruency in their minds even when there is no basis for it in fact. The folk wisdom is that groups will perform better when there is status congruency, because roles are better defined, making coordination easier.

Again, the folk wisdom may be questioned. The Melnick and Chemers (1974), Foeldesi (1976), and Klein and Christiansen (1966) studies cited earlier also examined status congruency, and no influence of its effect was found on performance. Adams (1953) measured status congruency, social liking, and performance of U.S. Air Force bomber crews, and found that in spite of a direct relationship between social liking and congruency, there was an unusual relationship between degree of congruency and performance, where performance increased with a small increase of congruency, and then fell off sharply as high levels of congruency were found. Adams offered no explanation for this finding; one may speculate that when status congruency is high, distinct social classes within a crew form which make communication and therefore coordination more difficult. This finding, which is over 30 years old, should be replicated using modern soldiers and modern statistical techniques.

Reddy (1975, p. 178) points out that questions of heterogeneity-homogeneity or congruence of member characteristics and compatibility of member characteristics are not the same, and describes an important distinction between the two:

> Homogeneity vs. heterogeneity implies dissimilarity of traits or variables, while compatible vs. incompatible implies non-complementary needs. Thus, while individuals may be homogeneous on a number of personality traits or variables, they may be quite incompatible in terms of their interpersonal needs.

A good example of homogeneous but incompatible groups is the homogeneous high-dominant group composition in the Lampkin (1972) study discussed in the previous subsection. The tendency for all group members to try to change each others' opinions was counterproductive for group functioning. To complicate matters, compatibility may involve homogeneity on some characteristics and heterogeneity on others.

One of the first researchers to develop a coherent theory and measurement system for compatibility was Schutz (1958), who hypothesized three interpersonal needs: inclusion, control, and affection. To measure an individual's desire to express behavior and desire to receive

behavior in each interpersonal need area, he developed a scale called the Fundamental Interpersonal Relationship Orientation-Behavior (FIRO-B). A sizable number of studies investigating the relationship between compatibility and group performance have used the FIRO-B scale.

Schutz used the FIRO-B scale to illustrate the relationship between compatibility and need for affection and productivity. He composed groups that were compatible—half of the groups preferred close, intimate relationships and half of the groups preferred to keep others at a distance—and groups that were incompatible (some members preferred close relationships and some preferred distant relationships). All groups were matched on general intelligence. The groups performed 14 tasks over a six-week period, including discussion tasks (choose a name for the group), modified chess-type games, and structure building. On all tasks, compatible groups outperformed incompatible groups. Interestingly, there was no difference in performance between the two types of compatible groups.

The research following Schutz on group compatibility and performance has been conducted in experimental and natural settings. Each setting will be considered in turn.

The experimental studies have used building, intellectual problem-solving, and creative writing tasks. Hewett, O'Brien, and Hornik (1974) and O'Brien, Hewett, and Hornik (1972) formed groups that were compatible or incompatible on all three interpersonal need areas on the FIRO-B scale. The groups were instructed to construct as many molecular models as possible within a 40-minute work period. The task organization was further divided into two conditions: collaborative, in which all group members were required to work together on all parts of the model, and noncollaborative, in which each group member had sole responsibility for building one model at a time. The group's score was the total number of correct connections between segments in the models constructed. In the O'Brien et al. (1972) study, compatible groups were more productive than incompatible groups in the collaborative condition, but were less productive than incompatible groups in noncollaborative condition. In the Hewett et al. (1974) study, in contrast, compatible groups were superior to incompatible groups in both conditions. Hewett et al. suggested that the somewhat conflicting results in the two studies may be due to the difference in leader power: in the 1972 study, an appointed leader was given instructions for the task and was directed to explain them to all other group members, whereas in the 1974 study, all group members received the instructions simultaneously. In the noncollaborative condition of the 1972 study, compatible groups may have spent more time discussing the project with the "leader" than doing the work.

Reddy and Byrnes (1972) also compared compatible and incompatible groups on a construction task. The group's task was to build a model of a man out of *Lego* blocks, copying a completed model. The measure of group performance was the time to completion. For each group, compatibility scores were computed separately for inclusion, control, and affection. Compatibility scores were then correlated with the group performance scores. Compatibility on control and affection were positively related to group performance, whereas compatibility on inclusion was not related to performance. Although reporting results separately by area of interpersonal need is informative and important, without also computing an overall compatibility index it is difficult to compare Reddy and Byrnes' results to those of Hewett et al. (1974) and O'Brien et al. (1972) reviewed above.

Instead of using a building task, Moos and Speisman (1962) gave dyads a pegboard task known as the "Tower of Hanoi" problem.[5] Groups were scored on the amount of time and number of moves needed to complete the task. Half of the groups were compatible on the basis of all three need areas, and half of the groups were incompatible. Compatible groups used significantly fewer moves than incompatible groups to complete the task, but the two kinds of groups did not differ in time to completion.

O'Brien and Ilgen (1968) used a similar design to that of Hewett et al. (1974) described above except that groups were instructed to write creative stories in response to TAT pictures. The stories were rated on plot originality, elaboration, plot structure, sentence structure, expressiveness, humor, and suspense. Unlike the other studies, compatibility did not relate to group performance.

Liddell and Slocum (1976) used a novel approach to form compatible, incompatible, and random groups on need for control in an intellectual task. Compatibility was not determined on the basis of group composition but on the congruence between group members' interpersonal needs and their appointed position of leadership in a task. In compatible groups, persons who expressed a need to exert control were placed in positions of influence, and persons who expressed a need for others to tell them what to do were assigned to peripheral positions. In incompatible groups, the need-position assignments were reversed, with need-to-control persons in peripheral positions and need-to-be controlled persons in influential positions. In random groups, members were assigned to positions at random. The task was to determine

[5]In this classical task, rings of different sizes arranged in a pyramid structure must be moved from one peg to a second, using a third peg as an intermediate location. The rules are that larger rings may never be placed above smaller ones, and only one ring may be moved at a time. For five rings, as in the present instance, it is possible to complete the task in 31 moves.

which two out of six geometric symbols matched. As hypothesized, compatible groups completed the symbol identification problems faster and with fewer errors than incompatible groups. The performance of random groups was in the middle.

The four studies conducted in a real setting related the compatibility of natural groups to their performance. The results were inconsistent and tended to conflict with those of experimental studies. Underwood and Krafft (1973) studied the work effectiveness of pairs of supervisors in manufacturing plants. Pairs of employees who worked together naturally were rated on two measures of interpersonal work effectiveness: (1) ratings by their supervisors, and (2) ratings on a simulated management task requiring group members to establish priorities and delegate work responsibilities. Unlike the studies reviewed in this section, Underwood and Krafft measured compatibility not on Schutz' (1958) three areas of interpersonal need but on Schutz' compatibility types: originator compatibility-behavior level, interchange compatibility-behavior level, originator compatibility-feeling level, and interchange compatibility-feeling level. Underwood and Krafft define these compatibility types:

> Originator compatibility compares the difference between an individual's express (transmission from self to others, e.g., the desire to lead others) and want (transmission from others to self, e.g., desire to be led by others) directions to the difference between another individual's express and want directions. For example, if the needs of one person are unbalanced favoring the express direction, a compatible other would have needs equally unbalanced favoring the want direction. Therefore, one would express as much as the other wants expressed. Interchange compatibility compares the overall intensity of one person's needs to the overall intensity of another's. The more similar the magnitudes of intensity [in any direction], the more compatible the individuals. (p. 90)

Each compatibility type combines the three interpersonal need areas (control, inclusion, affection). On the ratings of real interpersonal work effectiveness, only originator compatibility-feeling level was significantly related (positively) to effectiveness. On the simulated task, only interchange compatibility-behavior level was related (positively) to work effectiveness. Underwood and Krafft suggested that the absence of more significant correlations might be due to a restriction in the ranges of group compatibility. If so, this suggests that groups that are very high or very low on compatibility may not form naturally.

The second "natural" study took place in a natural setting, but the groups were composed by the investigator. Shalinsky (1969) investigated the performance of children at a summer camp on games (for

example, jigsaw puzzles, singing marathons) at the end of a three-week session. Children, aged nine to twelve, were assigned to cabins on the basis of need for affection. Compatible cabins were all high or all low on need for affection; incompatible groups had students with dissimilar needs for affection. As predicted, compatible groups won more games than incompatible groups. The result was explained on the basis of observed cooperative behavior in the compatible children during the camp. Again, compatibility rather than the location of the group on the need for affection scale was the differentiator among groups.

The third study produced findings conflicting with those of the above studies. Hill (1975) observed teams of systems analysts in the computer services department of a large oil company. The teams spent most of their time designing large computer systems. The outcome measure was performance on the most recently completed project. All teams were scored on a total compatibility index which reflected control, inclusion, and affection. Contrary to expectation, compatibility was *negatively* related to performance. Hill (p. 218) noted, however, that the task of designing computer systems in this study did not necessarily involve much interdependence: "members would go for several days without face-to-face contact as a group. Competitive impulses aroused by incompatibilities may thus have been channeled into individual task accomplishment. . . ." This hypothesis and the implicit suggestion that individual task accomplishment was related to higher performance need to be tested directly.

In the final study conducted in a natural setting, Hawley and Heinen (1979) examined groups of MBA students in a Business Administration program who worked on projects with host organizations (e.g., industrial settings) for a semester. Team performance was positively correlated with compatibility on need for inclusion and need for affection, but was negatively correlated with compatibility on need for control. The investigators did not suggest an explanation for the negative relationship, but instead emphasized the importance of maintaining separate measures of compatibility rather than pooling all measures into a single index.

It appears, then, that although there are some inconsistencies across studies, and serious reservations with respect to intact sports teams, compatible groups seem to be more productive than incompatible groups. This result was true for groups working on discussion tasks, intellectual games, several kinds of building projects, creative writing, symbol-matching, management problems in industry, and children's games.

SUMMARY

The research relating group structure to group performance has primarily examined the effects of heterogeneity of individual characteristics of ability and personality on group performance. For general and (in particular) specific abilities, it was found that for tasks of a coactive nature, homogeneous groups performed better, whereas on tasks of an interactive nature, heterogeneous groups were superior. This finding recalls the earlier conclusion with respect to individuals' characteristics that the nature of the task influenced the effects of ability on performance. It is likely that the underlying causes of the two findings are similar—that in coactive tasks, high-ability individuals and low-ability individuals can have major influence on outcome, while in interactive tasks, the group can adapt to use its talents in optimal ways. Thus, for coactive tasks, homogeneous groups will tend to be free of low-ability members, and therefore perform better than heterogeneous groups, while in interactive groups, the high-ability members in the heterogeneous groups will be effectively utilized. Although the studies examining personality and cognitive compatibility are fairly consistent, suggesting that compatible groups outperform incompatible groups, there is evidence that ability and personality homogeneity interact such that task demands appear to shape the relationship between group composition and performance. Therefore, as a basis for generalization it is best to use those studies that have used tasks with relevant characteristics. In particular, for military tasks with specific requirements for general abilities, specialized skills, and interdependence among group members, it is necessary to delineate these specific requirements before we can know how best to structure a task unit to maximize productivity.

VI. GROUP PROCESSES

Group processes are characteristics of functioning groups that arise after a group's formation, and do not exist independently as a group itself. Where the group structure might be thought of as a combination of individual contributions, group process is more than the sum of the group parts. The particular aspect of group process of present interest might be characterized as the social psychological climate of the unit, or the nature of the individuals' perceptions of their interpersonal and environmental work-place (Gavin and Howe, 1975). Various attempts have been made to characterize this climate. Sventaitskiy (1973) informally assessed the major factors of the social psychological climate of the work-place to be (1) the extent of interest in workers' tasks, (2) leadership style, (3) level of interpersonal compatibility among workers, and (4) the predominant economic system in the work-place (capitalist vs. socialist for this Soviet author). Jones and James (1979), in a study involving over 5000 subjects and extensive questionnaires, identified six dimensions of psychological climate: (1) the challenge posed by the job, (2) leadership style, (3) interpersonal compatibility among workers, (4) professional and organizational esprit, (5) conflict and ambiguity in the work environment, and (6) the demandingness of the task. It is interesting to note that the first three factors of each study coincide; Sventaitskiy's last factor is clearly political, whereas the statistical power of the Jones and Jones study enabled it to uncover additional factors.

Our focus will be on those dimensions of social psychological climate that are inherent in the working group rather than the task itself. This means that Jones and James' third and fourth factors, covering the interpersonal and person to group relationships of the performing unit will be of most interest. Throughout, we will examine the climate from the individual's point of view rather than from the organization's (Gavin and Howe, 1975). The major topics of discussion are group cohesion, or the extent to which there are forces drawing the group together, and attraction, or the liking of the members of the group for each other. These and other aspects of group process are examined below.

Two studies illustrate the types of issues raised under the rubric of group processes. Goodacre (1953) interviewed the 13 best and 13 worst rifle teams from 63 infantry squads, to ascertain the dimensions that differentiated them. The interviews were open-ended, oriented toward

five dimensions of interest, and chi-square statistics were done on the coded answers on an item-by-item basis. Of the five dimensions of team stability, potency, liking, intimacy, and stratification, three showed differences between the best and worst squads. Good squads were more potent in that the group was perceived as more important in the members' lives; had better liking for each other; and had a clearer separation of leader and subordinate roles than did poor groups. The stability (turbulence) factor was not significant, in keeping with the studies reported above, nor was the intimacy factor.

Magen (1980) reports a remarkably successful intervention to improve sports performance. A leading Israeli soccer team that had excellent personnel but was losing asked Magen to conduct a series of encounter group sessions with the team coaches and members. In the group, Magen focused on members' awareness of their own responsibility for the team's performance, and obtained from each a public commitment to be different in one specific way in order to improve the team's performance. The result was an immediate reversal of the losing pattern, leading to a sequence of wins, including a victory over the top-ranked team in the country. Magen argues that this is a demonstration, albeit not a proof, of an argument that solidarity improves group performance. Below, we will examine this argument in some detail.

GROUP COHESIVENESS

Group cohesiveness is a major focus of interest, especially among sports psychologists (see Hare, 1976; Lott and Lott, 1965, for general reviews). The folk wisdom is that group cohesiveness leads to improved group performance, but that wisdom has been in the process of qualification for some time. One of the earliest studies of group cohesiveness assembled two groups of carpenters and bricklayers in a large housing project on the outskirts of Chicago: one group was composed of workers who preferred to work with each other; the other group was composed as usual, without regard to preferred coworker (Van Zelst, 1952). Moreover, the two groups were matched on previous performance. At the end of the three-month project, the experimental group had a lower turnover rate and lower labor and material costs than the control group, suggesting that the experimental group finished subtasks in less time than the control group. These results may have been due to a "Hawthorne effect," where greater attention is paid to the experimental group.

Hagstrom and Selvin (1965) employed a questionnaire to analyze the cohesiveness of 20 groups of women living in sororities and dormitories, and found two dimensions of cohesiveness. The first was labelled *social satisfaction*, and indicated the extent to which the group provided the individual's needs. This instrumental cohesiveness was shown to be related to time spent dating, whether or not the individual voted in student elections, and other concrete matters. The second factor was labelled *sociometric cohesion*, and measured affective coherence, or liking for being in the group. This cohesiveness increased as the proportions increased of an individual's best friends who were in the group, the proportion who sought advice within the group, and subjective feelings of closeness. Although Hagstrom and Selvin did not specifically examine productivity, they conjectured,

> in strongly task-oriented groups, group effectiveness may be a major determinant of attractiveness, and effectiveness may be hindered by too high a degree of sociometric cohesiveness. (p. 40)

In other words, instrumental cohesiveness could be promotive of group productivity, whereas affective cohesiveness could hinder productivity. The explanation of the hindrance of affective cohesiveness for task performance was that cohesive groups direct much of their efforts toward integration of group members, rather than to the task (Fiedler, 1953, 1954; Stogdill, 1974; Horsfall and Arensberg, 1949; Bass, 1960; Feldman, 1969; Zander, 1969).

The Hagstrom and Selvin paper has been cited in several sports psychology studies examining the influence of team cohesion on success of sports teams. These studies have largely focused on affective cohesion in some form, and cover many sports in a number of nations. Ball and Carron (1976) examined ice hockey, Bird et al. (1980), Gruber and Gray (1981), Klein and Christiansen (1966), Martens and Peterson (1971), and Widmeyer (1977; Widmeyer and Martens, 1978) all studied basketball, Bird (1977) and Voe and Brinkman (1967) studied volleyball, Landers and Crum (1971) studied baseball, Landers and Luschen (1974) examined bowling, Foeldesi (1976) studied rowing, and McGrath (1962) and Myers (1962) studied rifle teams. Summaries of this work have been written by Carron (1980), Landers, Brawley, and Landers (1981), and Straub (1975). The history of this research tradition is one of increasing methodological sophistication, as care comes to be taken concerning the time of measurement (when in the season are cohesiveness measures taken), statistical techniques (moving from simple univariate breakdowns to complex multivariate models), use of control groups and even artificially constructed teams, and the definition of cohesiveness (simple scales to more established measures). It has been

shown that affective cohesiveness is associated with success in sports efforts involving divisible tasks such as bowling, rifle marksmanship, and to some extent baseball, but that affective cohesiveness can hinder performance in sports requiring close task-oriented coordination, such as basketball, volleyball, and ice hockey. An interpretation is that when there is a strong degree of affective cohesiveness in a team, energies are spent keeping the team together, which means that there is less critical appraisal of performance, and more expressions of unconditional positive regard. In less affectively cohesive groups, task-orientation is more predominant, and players receive social reinforcement contingent on the quality of performance.

Examinations of instrumental cohesiveness have generally shown that successful teams are more cohesive. However, as the time that measurement is taken is more carefully considered, the causal direction of success inducing cohesiveness is becoming increasingly more likely than cohesiveness leading to success (Landers et al., 1981). For example, Bird, Foster, and Maruyama (1980), in one of the most methodologically sophisticated studies done in the area, demonstrated a relationship between instrumental forms of cohesiveness and success that only emerged at the end of the season.

Additional evidence that performance causes cohesiveness comes from a study by Bakeman and Helmreich (1975), who did an intensive study of marine scientists living together underwater for a period of time. As part of this study, tests of productivity based on evaluations of the scientists' publications and tests of group cohesiveness based on full-time observation of the enclosed environment were made at different times. The order of cohesiveness and productivity were tested in a cross-lagged panel analysis, from which it was concluded that performance precedes cohesion. This particular study, examining as it does an environment and task completely different than that of a sports team, considerably buttresses the weight of evidence of the earlier studies.

INTERPERSONAL ATTRACTION

Interpersonal attraction, which could be viewed as an element of cohesiveness, has been singled out for particular attention. McGrath (1962) examined the relationship between positive interpersonal relations and effectiveness in ROTC rifle teams. He began by having 60 three-man rifle teams rate members on warmth and attentiveness at the end of a rifle tournament. Using the behavior of the *raters* (rather than impressions of ratees), he created new teams whose members were

either interpersonally oriented (saw others as warm), not interpersonally oriented (did not see others as warm), or mixed. Not counting the mixed group, 35 reconstituted teams were available for further testing. A principal result was that the non-interpersonally oriented teams continued to improve in the training week following the tournament, whereas the interpersonally oriented teams did not. In the non-interpersonally oriented teams, performance was related to individuals' social adjustment and satisfaction with the task, while in the interpersonally oriented group, this did not obtain. Overall, McGrath found that the reconstitution into new teams facilitated the performance of the non-interpersonally oriented group, but at a cost of member attractiveness. In the interpersonally oriented group, performance did not improve, and group adjustment was based on mutual liking rather than task performance.

Goodacre (1951) asked sociometric questions of 14 rifle squads from an Army infantry company. Members were asked to name the people they would associate with in nonmilitary, garrison area, and field/tactical area recreational time. The number of within-squad choices on each dimension was correlated with performance on a six-hour field simulation exercise. The correlations were all positive and significant. The better the group performance, the more team members were likely to associate with each other recreationally. Goodacre attempted to retest his squads, but at the time of his attempted follow-up, squads had been either broken up or shipped to Korea, and the project was abandoned.

Berkowitz (1956) examined how patterns of perceived similarity of attitudes and crew liking related to aircrew effectiveness in war combat in Korea. His subjects were 11-man crews flying B29 missions over Korea in 1953; performance was based on ratings of superiors and on the percentage of missions carried out as planned. Crew members' attitudes toward their jobs were assessed. In addition, for each attitude item, members were asked how many of their fellow crew members would agree with their own judgment. Finally, sociometric measures were employed to assess the degree to which crew members liked each other. Thus, the major predictors were attitudes toward their task (labelled "motivation" by Berkowitz), agreement among members on motivation, understanding by members of each others' motivations, and liking. Analysis of these measures showed no clear-cut relationship of these variables to performance. Instead, when there was high liking, performance was related to the crew's mean level of motivation for the task, but when there was not high liking, performance was a function of the degree of understanding members had of each others' positions. It appears that when members agree, they understand and

like each other, and perform in accordance with the group norm, as reflected by the mean motivation level. But when they don't like each other, then understanding permits them to focus on the task at hand and "get the job done" without recourse to the collective motivation of the group. Put another way, for groups with high affiliation, the affiliative cohesion results in performance in accordance with the group norm, but for groups without high affiliation, a task orientation can be adopted if they understand each others' motivation, but not if they don't have such an understanding. Berkowitz' separation of agreement and understanding is a fruitful way of conceptualizing interpersonal relationships that has been successfully employed in social psychological arenas other than group performance, and should be adopted with greater frequency in future research efforts.

SUMMARY

The literature on group cohesion indicates that deliberately inducing social cohesion, either of the instrumental or affective type, will not significantly improve performance in the interactive coordinated tasks that typify military units. Indeed, too much affective cohesion might interfere with the critical appraisal of performance that is needed to maintain quality, and instrumental cohesion is perhaps generated as a consequence, rather than a cause of group productivity.

An argument might be made for carefully monitoring the extent of affective cohesion. The socially cohesive groups examined appear to manifest some of the pathologies of "groupthink" (Janis, 1983), a systemic concern for solidarity that yields sometimes severe decrements in the quality of group decisionmaking. It might be worthwhile to see if Janis' proposed groupthink preventive measures can be extrapolated to nondecision tasks.

VII. TRAINING TECHNIQUES

The final topic to be covered in this literature review is team training. The question of when team training is better than individual training, and how team training should proceed, are questions that have been important to the military for some time. There have been recent surveys such as analytic reviews by Hall and Rizzo (1975) and Thorndyke and Weiner (1980) for the Navy, and a descriptive review by Dyer et al. (1980) for the Army. Goldin and Thorndyke (1980) report on a workshop at The Rand Corporation devoted to the topic of team training. We will summarize the results of these reviews, and cover several additional aspects of training techniques not discussed elsewhere.

FEEDBACK ABOUT PERFORMANCE

Feedback is perhaps the most relevant aspect of team training in this review. The two topics most important for the prediction of team performance are (1) feedback vs. no feedback and (2) the aggregate level at which feedback is directed—individual or group. Each topic is considered in turn.

Nadler (1979) reviewed the literature comparing group performance under conditions of feedback and no feedback. The majority of studies making this comparison have used feedback about the group's performance rather than feedback about the performance of individual group members. Not surprisingly, groups that received feedback about their performance tended to improve over time whereas groups receiving no feedback maintained the same level of performance or declined over time (see, for example, Bowen and Siegel, 1973; Cook, 1968; Kim and Hammer, 1976; Pryer and Bass, 1959; Walter and Miles, 1972; Weber, 1971, 1972; and Zander and Medow, 1965).

Not all studies have found feedback to be effective, however. Ellsworth (1973) and Spoelders-Claes (1973) found no difference between feedback and no-feedback conditions. Glaser and Klaus (cited in Zander, 1971) describe a situation in which feedback can actually be *detrimental* to group performance. This situation arises when group success is contingent upon the performance of a single group member or a subset of group members, rather than all group members. When feedback pertains only to the performance of the group, and the group

is successful, members performing moderately well or poorly may assume that they are performing successfully. Over time, their performance is expected to deteriorate because they make no effort to improve, and in turn the performance of the group will worsen.

In the scenario described above, feedback about the group's performance may be detrimental to group functioning. This problem raises the question of the best level at which to target feedback. A considerable body of research has investigated this question by comparing feedback about the group's success with feedback to individual members about their own performance in the group. This research consistently reports that feedback about individual performance has a much greater impact on group performance than feedback about only the group's performance (see reviews by Meister, 1976; Nadler, 1979; Zander, 1971; as well as studies by Berkowitz and Levy, 1956; Rosenberg and Hall, 1968; Smith, 1972; Stone, 1971).

Furthermore, a combination of group and individual feedback is often more effective than any one type of feedback given alone (see, for example, Zajonc, 1962; Zander and Wolfe, 1964). In Zajonc's study, group members were required to press a button as soon as a specific stimulus light appeared. The performance measure was reaction time, and the group scored a point whenever a prespecified number of group members reacted in a certain amount of time. There were two feedback conditions: feedback about the group's performance only, and combined feedback about an individual's performance, the performance of other group members, and performance of the group as a whole. Groups receiving combined feedback showed substantial improvement over time; groups receiving only group feedback showed little improvement.

The drawback of Zajonc's study is that separate comparisons could not be made about different kinds of feedback. Zander and Wolfe designed their study to allow such comparisons to be made. The four feedback conditions were (1) group only, the sum of scores of all members, (2) individual only, the separate scores of each member, (3) group and individual, and (4) no feedback. The five-person groups were members of district coordinating committees in a large utility company. Their task was to predict which two out of a total of four events would occur on each trial. Group members could combine their resources in any way they chose. Group performance improved only under the combined feedback condition. Neither group nor individual feedback was effective when given alone.

After reviewing many studies comparing different kinds of feedback, Nadler (1979) proposed a model to explain when certain kinds of feedback (group vs. individual vs. combined) would be effective. In his

model, group feedback is expected to have the greatest impact on group performance when the task requires interdependence among group members and where group members have individual responsibilities. Individual feedback is expected to have the greatest impact on group performance when group performance is merely the sum of individual performances.[1] Given the research results described above, however, it seems safe to expect that combined feedback about group and individual performance can never be detrimental.

TRAINING THROUGH SIMULATION

The Army has conducted several research projects on the effectiveness of different forms of simulated engagement as a training tool. The near-universal opinion is that such models as REALTRAIN (Meliza et al., 1979; Scott, Meliza, Hardy, and Banks, 1979; Scott, Meliza, Hardy, Banks, and Word, 1979) and COTEAM (Modlin, 1979) have proven more effective training techniques than the ARTEP models they replaced. Sulsen (1980) reports that repeated engagement simulation exercises improved individual and collective performance of rifle squads, and Root et al. (1979) report the great superiority of engagement simulation techniques in training and evaluating units because such a technique is the only way to train for the experience of interactivity within a training unit and of reacting to enemy strategizing. Although some of these studies suffer from small sample sizes and decidedly nonrandom assignment of personnel to condition, the weight of evidence of each of the studies is so strong that one is forced to accept Root et al.'s conclusion. Indeed, as Meliza et al. (1979) note, REALTRAIN has been adopted as the preferred method of combat unit training.

The usefulness of nonengagement simulation training is also promising. Miller and Bachta (1978) report that the Dunn-Kempf tactical board game has taught command and control leader training. Using this game, commanders-in-training learn how to establish priorities, use communications effectively, and maintain adaptability, all without having to undergo the expense of using enlisted men to actually carry out tactics. How such a board game transfers to actual team performance, where the range of possible outcomes is greater and uncertainties magnified, is not well-established; a conservative judgment is that such board games might supplement but not replace engagement simulation training for commanders.

[1]This distinction recalls the one between interactive and coactive tasks made earlier.

MOTIVATION IN TRAINING

Bird (1978) and Zander (1978) both point out that there is a difference between the motivations of the individual members of a group and the motivations of the group itself. Both suggest that care should be taken to motivate individuals within a group in order to guide the group toward success. Zander distinguishes between

- Supportive training, where trainees are given support that is not contingent on performance,
- Reinforcement training, where trainees are given rewards for good performance, and
- Pride-in-performance training, in which an internal motivation to achieve success is aroused.

Bird recommends the third of these training techniques and offers a means of carrying it out, which includes setting specific goals that are achievable by the group, moving toward competence in small increments, and offering the group as a whole as many successful experiences as possible. This success should lead to a greater group cohesion, which in turn will promote interactiveness. This model has not been tested to date, but is worthy of research.

TEAM-ORIENTED TRAINING

Several studies have examined training that specifically focuses on the team rather than its individuals. Among the topics falling under this rubric are clarity in job definition, communications among members, and evaluation of present military training practices.

Cory et al. (1979) have questioned the level of detail of job definitions as it impinges on training effectiveness. Their research, which was not empirically based, examined how it might be possible to group military jobs into units smaller than an MOS but more general than a specific task. Such groups of tasks would have representative descriptions and training materials prepared for them. This would, Cory et al. argue, lead to more efficient and effective training. Their conjecture certainly has merit, and should be subjected to empirical testing.

Siegel and Federman (1973) examined the communications of anti-submarine helicopter crews in order to identify what characteristics of communications differentiated between good and poor performance. Performance was defined as miss distance in an anti-submarine warfare (ASW) exercise. They obtained 25 measures of communication, of which 18 were significant predictors of performance. These 18

measures were analyzed in two separate data samples, from which emerged three common factors:

1. Leadership control, or the extent to which opinions of crew members were allowed to emerge,
2. Probabilistic structure, or the explicit weighing of probabilities, and
3. Evaluative interchange, or the exchange of ideas, proposals, and data among crew members.

Although these factors were not compressed into composites and subjected to multiple regression, and although the cutoff of the factor loadings for assigning different communication variables to the three factors was not as high as common wisdom would recommend, the three factors do have some face validity. After the factors had been identified, a group of 32 crews were randomly divided into 16 controls who received normal training and 16 crews who were trained to employ the three factors in their communications. This training did not, however, lead to statistically significant differences between groups on number of hits, success on a paper and pencil test, or a mean miss distance on torpedo firing tests. The experimental group did have a lower mean miss distance for the firing tests, and a much lower standard deviation of miss distance (which was not tested against the control group in the paper), so it is possible that reanalyses adjusting for this nonhomogeneity of variance might yield statistical significance. However, overall, this study cannot be used to recommend specific changes in communication training to improve team performance.

Dyer et al. (1980) present the beginning of an Army effort to understand teams and team training. Earlier, we reviewed their definitions of team and distinction of types of team. Here, we discuss their work on team training. As part of the questionnaire distributed to 140 units through FORSCOM, team leaders were queried about the adequacy of specific parts of training. The ideal amount of training was compared to actual training received for different types, and several deficiencies, particularly with regard to combat units, were noted. Although special school training occurs on the average less than once a year, it was thought to be desirable several times a year. Field training, which takes place several times a year, should be performed close to monthly. Training devices, used almost once a month, should be used several times a month. And on-the-job training should occur weekly instead of the several times a month it was found to occur. In general, though, leaders were moderately satisfied with training (mean of 2.3, where 1 = completely satisfied and 5 = completely dissatisfied). The greatest

complaint was that there was not enough time for training, with the inadequacy of scheduling of what time did exist also being noted. Insufficient training was the main complaint of all operational problems noted, although this was not seen as a critical complaint by the team leaders surveyed.[2]

Hall and Rizzo (1975) assessed the state of knowledge regarding Navy tactical team training. Their findings on team definition and evaluation were discussed earlier. They summarize the state of the art regarding team training as follows:

> While everyone professes intuitively to be able to recognize a good team—the "I'll know it when I see it" phenomenon—no one seems to be able to articulate its dimensions with sufficient clarity to permit the development of training procedures for producing it. (p. 15)

They divide the question of team training into consideration of (1) the extent to which a team is a collection of individuals vs. a collectivity which must be uniquely trained, (2) how to train coordination, and (3) how to train groups to be cohesive. Within the framework of this breakdown, their review of the literature on team training resulted in several concrete recommendations. The recommendations dealt largely with developing concrete objectives to achieve through training, composing a training environment that was conducive to generalization to the tasks being trained, and concentrating more on high-quality individual performance than on training at the level of the performing unit. For example, it is suggested that performance feedback is more effective if it is at the individual member level rather than a global team feedback, as individuals are often unaware of how their specific performance contributes to the unit as a whole. Hall and Rizzo's (1975) survey, as well as Rizzo's (1980) summary of its main points, are documents worth perusal on their own; full justice cannot be provided them in a literature summary.

Thorndyke and Weiner (1980) designed a research program to improve training and performance of Navy teams. The thrust of their approach was oriented toward team decisionmaking; very little effort was directed toward team performance requiring coordination of perceptual-motor skills such as might be used in armor or infantry units in the Army. Rather, they were concerned with high-technology solutions to questions of high-technology utilization. For example, they advocate intensive research on decisionmaking teams through development of highly computerized experimental laboratory facilities. Their envisaged research center includes experts in artificial

[2]Commissioned officers and consultants have taken a dimmer view of training; see Madden (1981).

intelligence, software and systems design, and cognitive psychology, as well as expertise in human factors, social psychology, and simulation and gaming. The recommendations made in this document are abstract, and not directed at immediate improvement of team performance, but rather at how one could learn what is needed to improve team performance. These recommendations are especially welcome, given the generally inadequate state of knowledge about team performance that this review has demonstrated.

Finally, Goldin and Thorndyke (1980) summarize a three-day workshop held at The Rand Corporation under the sponsorship of the Office of Naval Research that was devoted to improving team performance. Although some of the papers from that workshop have been reviewed earlier in this report, it is worthwhile to reiterate some of their conclusions with regard to team training. First, they find organizational constraints on team training effectiveness. For example, a lack of standardization can hamper training efforts, as trainees learn on one system, and then are tested on a second system, and finally perform the job on a third system. The transfer of what is presumed to be the same skill from one system to another may not be direct, and efficiency is reduced. Also, the organizational climate may not be conducive to effective training, as trainees compete for scarce promotion slots and limited rewards. Efforts are then directed at individual recognition above that of fellow team members rather than either individual or group productivity.

Second, problems are noted with the team training practices themselves. Goals of training are typically too abstract, rather than phrased in terms of specific procedural skills or objective performance criteria. The objectives of individual proficiency and group coordination are often mixed together in the same training task, so that it is difficult to sort out the two for the individual being trained, and neither objective is fulfilled. The lack of clear feedback on individual performance and the lack of standardized equipment noted above both contribute to this problem. The conference concluded that too much training is aimed at teams as units rather than at individuals as units; this approach loses sight of a goal of individuals within teams being seen as interchangeable parts bringing their own particular expertise to merge with that of the other members to yield effective group performance.

The recommendations of the conference were for research rather than specific prescriptions for change. Among the areas recommended for study were:

1. How to train procedural flexibility. Modern military teams are faced with uncertain, complex environments for which no simple algorithmic rule of behavior is likely to be successful. Instead, teams must be flexible and adaptable to the environment and strategic moves of the opponent.
2. Team structure. It is not clear what is the best type of team structure for any given team task. A consensus was that the task was important in determining the optimal structure, and that the relationship between informal and formal structures required investigation. Centralization can be either beneficial or harmful, depending on the particular task.
3. Team communication. All feel that communication is critical, yet it is not known how to design communication channels that maximize team performance.
4. Organizational environment. Factors external to the team influence the team's effectiveness. These factors may be under at least the partial control of the organization, and therefore should be understood so that they may be optimally manipulated.

SUMMARY

The recommendations of the Rand conference noted above also summarize much of what is known regarding team training techniques. Procedural flexibility is a difficult concept to operationally define, much less incorporate into training; perhaps it might best be instilled through pride-in-performance training, where motivation to achieve success is inculcated and team members feel free to interact with each other and to take initiatives to modify specific jobs when necessary in the service of success. The team structure during training refers not only to the tasks for which members are trained, but also to the structure of the training itself. The literature indicates that feedback on both the individual and group levels is safer than either type of feedback alone or no feedback, in that task structures of different types can be more easily learned. This is in effect purchasing training insurance, at the small cost of some perhaps superfluous feedback. In addition to considerations of feedback, the definition of individuals' tasks within the group effort should be carefully considered, so that each member knows what his own role is and how he interacts with the other members. Training should also include communication, so that members know how to communicate effectively, and can choose the best communication channels in the face of diverse problems. Finally,

the organizational environment, including how the various members of a team interact and how the team will function in the "real world," must be represented in training; this appears to be done well by sophisticated simulation techniques.

VIII. CONCLUSIONS

In this review, we have examined the nature of unit performance and searched for predictors of quality performance. Here, we shall recapitulate what has been learned as a result of the review, both in terms of what is known and of what directions future research should take.

Definitions. In defining unit performance, there is a problem of the level of analysis, or what object should be the focus of scrutiny. For some questions, the small group comprising the unit is the focus, whereas for others, it is the individuals in the group who must be scrutinized. An orderly approach to this problem, using a framework such as Living Systems Theory, is recommended.

Similarly, the nature of the group task to be performed merits careful attention. In particular, the degree of interaction that is required among the group members is a critical dimension to be considered. Second, the degree to which group members have distinct roles as opposed to interchangable positions is important. The military tasks of today's interest may be characterized as having a common goal to which all members aspire, having a division of labor among the group members, and requiring coordinated, interactive effort among the members.

Finally, the definition of performance must be attended to. Granted, any measure of military performance in peacetime is an imperfect surrogate for actual behavior in combat. But performance has too often been assessed by global judgments by supervisors, superior officers, or teachers. These subjective judgments are not reliable over different times, environments, or raters, and may be of questionable validity as well. Instead, more objective measures of performance, in the form of composites of relatively straightforward judgments of small behavior segments, are urged. In almost every comparison of evaluations, objective measures have been shown to be superior to subjective ones.

Characteristics of Individuals. The individual characteristics studied here were general ability, ability to perform specific tasks, motivation, and personality. The effects of the individual characteristics of the extreme members of the group (e.g., most able, least able) as well as the average group member were examined.

Many of the studies surveyed demonstrated substantial correlations between member ability (both general ability and ability to perform specific tasks) and unit performance. However, those substantial

correlations did not obtain for all of the studies, and in particular were not found in studies of performance in military tasks. When all of the studies were considered together, it became apparent that the relationship between ability and group performance depended on the nature of the task. For tasks requiring contributions by all group members, the proficiency of the least-able members and the average proficiency were important predictors, whereas when the task could be completed successfully by dint of superior performance by only some group members, then the most-able member's ability was important. Among the implications of these findings are that (1) studies of ability must take into account the nature of the interrelationships of individuals' tasks within the group; (2) the different ways that groups face tasks should be further investigated and compared to optimal performance models; and (3) performance as a function of member ability for different member roles should be investigated.

Only weak or inconclusive findings for the effects of personality characteristics of group members on performance were found. It does not appear promising at the present time to use personality measures in determining group composition. On the other hand, the motivation of individual members does affect the performance of the group. The research shows, however, that supervisors' anticipations of what motivates good performance may not necessarily be what actually motivates performance. It is suggested that learning what unit members wish to obtain from their tasks is useful in constructing incentive structures that will motivate superior performance.

Leadership. Studies of leadership did not provide any concrete suggestions for how leaders can be selected to improve unit performance. These studies did point out that there are inherent methodological problems in essaying such a task, because leader quality tends to be defined in terms of the performance of the leader's subordinates. The positive findings of leadership research all indicate that the question of leader effects is complex, depending on an interplay of style of leadership, task environment, interpersonal relations among group members, and task structure. At present, it is difficult to obtain a simple answer to whether a particular leader manipulation would or would not improve group performance.

Group Structure. This section examined the same major predictors as the section on individual characteristics, but looked at the effects of the heterogeneity of individuals' positions on those characteristics instead of the positions of specific or typical individuals. Indeed, on several occasions, the same studies were discussed in both sections. The findings of the group section paralleled those of the earlier one, in that heterogeneous groups performed better on tasks of an

interactive nature where members of the group were substitutable one for the other, but that homogeneous groups were superior when the tasks called for member specialization.

There was a consistent finding that groups compatible (homogeneous) with respect to several personality and cognitive style measures performed better than groups heterogeneous on those measures. This finding was tempered by consideration of the nature of the task and the ability of the group members, such that task demands appear to shape the relationship between group composition and performance. To best structure a unit to maximize productivity, it is necessary to know the degree of interdependence and skill levels required by the task. For more complicated, interdependent tasks, homogeneous groups are probably preferable, whereas for simpler, less interdependent tasks, heterogeneous groups may be constructed. These findings are, however, based on a small number of studies, and further verification of the findings should precede a decision to undertake the expense of obtaining personality and cognitive style measures and employing them in personnel decisions.

Group Processes. The literature on group cohesion indicates that deliberately inducing social cohesion, either of the instrumental (task-oriented) or affective (solidarity-oriented) type, will not significantly improve performance in the interactive coordinated behaviors that typify military tasks. Too much affective cohesion might interfere with the critical appraisal of performance that is needed to maintain quality output, as members become concerned with supporting each other and raising group morale instead of concentrating on the task at hand. While raising instrumental cohesion might theoretically be of benefit, there are no studies demonstrating this phenomenon, and indeed there is some indication that the association of productivity with high instrumental cohesion is due to productivity causing cohesion rather than the other way around.

Team Training Techniques. The research on team training techniques generally supported the present advances being made in military training. The importance of feedback, both on the level of individual members' performance and on the level of unit performance, cannot be overemphasized. Although it is not entirely clear when individual feedback is more important than group feedback, it is probably a good insurance policy to incorporate both into any training program. Inducing team motivation was touched on in the section on individual characteristics. A motivational set that induces members to have pride in performance as opposed to doing the job to obtain specific rewards appears to be the more promising for producing quality performance. Additionally, when motivation is bonded to group performance,

members will be freer to take initiatives when necessary, and interaction among members will be more adaptable to circumstances. Team training should be considered superior to individual training in tasks where each member must know his own role and how that role interacts with the roles of other team members. This team training should include consideration of how to achieve efficient and effective communication, and how the team as a whole fits into the external environment of which it is a part. Modern sophisticated simulation exercises appear to be good training tools to effect these objectives.

Appendix

THE RELIABILITY AND VALIDITY OF UNIT PERFORMANCE MEASUREMENTS[1]

In this Appendix we present generalizability (G) theory (Cronbach, Gleser, Nanda, and Rajaratnam, 1972), an analytic technique for examining the reliability of measures of unit performance, and then briefly treat validity issues. As a heuristic for demonstrating the application of G theory to reliability estimation, data on tank crew performance were used. These data provided a realistic context for specifying the requirements of a measurement model.

A measurement model should be capable of estimating the magnitude of error introduced into the measurement as a consequence of using different observers, occasions, and missions (Cronbach et al., 1972; for a recent review, see Shavelson and Webb, 1981). For example, as a measure of unit performance, decisionmakers should be willing to accept a single score provided by one observer, on one occasion, for a particular mission (and so forth) as representative of any number of possible scores that the unit might have obtained with different observers, on different occasions, and for a variety of missions. A measurement model should also be capable of capturing another obvious and critical feature of unit performance data—that they are multilevel (e.g., Burstein, 1980). Organizationally, crews are formed by individuals, platoons by crews, companies by platoons, and so on. To a greater or lesser extent, organizational variables affect unit performance and a measurement model should incorporate them.

This Appendix is divided into three parts. First, we outline a theory for examining the reliability of unit performance measurements. We then specify and statistically evaluate alternative measurement models examining the multilevel unit performance measurements. Finally, we discuss issues pertaining to the validity of unit performance measurements.

---

[1]This Appendix was primarily written by Richard J. Shavelson.

SKETCH OF GENERALIZABILITY THEORY

Reliable and valid measurements of unit performance are extremely difficult and expensive to achieve. To minimize cost, these measurements are often based on judgments of single observers who have more than one function to perform, and must evaluate multiple events that occur simultaneously. They are taken under a variety of conditions (e.g., no two terrains for carrying out a mission are exactly the same; no single opponent trained for a simulated exercise behaves exactly the same upon repetition of the exercise). And they are taken on different days, time of day, and so on.

Nevertheless, scores such as those assigned to tank crews in simulated missions (e.g., ARTEP Table VIII exercises) are interpreted as *characteristic* of the unit. Decisionmakers interpret these scores as interchangeable with scores that would have been obtained on a multitude of different terrains with any of a large number of different observers, being carried out against any of a large number of opponents, on any day. In other words, decisionmakers are willing to generalize a tank crew's Table VIII score over terrains, observers, opponents, days, and so forth. Ideally, the decisionmaker would like to know the crew's average score over all possible observers, terrains, and opponents. The issue of the reliability of a measurement, then, resolves itself into the question of *how dependable is the generalization from a single score to the average score the crew would have earned over all possible measures of its performance?*

Reliability thus refers to the *generalizability* or *dependability* of scores (Cronbach et al., 1972). As the number of facets entering into a measurement (such as observers, terrains, and occasions) increases, the possibility of introducing error into the measurement increases. As error increases, the generalization from a single score to the tank crew's average score may become increasingly less reliable. That is, as the number of *facets* of a measurement increases, the number of potential sources of error in a measurement increases. Increasing error creates increasingly unreliable measurements.

A measurement model should estimate the magnitude of error introduced into a measure of unit performance by each facet. Moreover, it should provide information on how to reduce that error in the most cost-effective way. Generalizability theory provides the basis for accomplishing this.

The facets of a measure of unit performance (e.g., observers, terrains, missions) define the *universe* to which a decisionmaker wishes to generalize. A *universe score* is the datum the decisionmaker ideally would like to know but must infer from a sample, i.e., from an observed

score. The universe score is defined as the unit's mean score over all possible observations in the universe. The difference between a unit's universe score and its observed score reflects error in generalization. Unit k's observed score, then, may be decomposed into a component for the universe score, μ_k, and one or more error components.

We illustrate this decomposition for the simplest measurement case, units and one measurement facet, say observers (o). The presentation readily generalizes to more complex designs. We assume, for simplicity, that the same team of observers evaluates all units. Hence, units and observers are crossed, and we represent the measurement design as: $K \times O$. In the $K \times O$ design with generalization over all admissible observers taken from an indefinitely large universe, the score assigned to a particular unit (k) by a particular observer (o) is

$$\begin{aligned} X_{ko} &= \mu && \text{grand mean} \\ &+ \mu_k - \mu && \text{unit effect} \\ &+ \mu_o - \mu && \text{observer effect} \\ &+ X_{ko} - \mu_k - \mu_o + \mu && \text{residual} \end{aligned} \qquad \text{(A.1)}$$

Except for the grand mean, each score component has a distribution. Considering all units in the population, there is a distribution of $\mu_k - \mu$ with mean zero and variance $\sum(\mu_k - \mu)^2 = \sigma^2_K$, which is called the universe-score variance and represents consistent error-free variation between units. Similarly, the component for observers has mean zero and variance $\sum(\mu_o - \mu)^2 = \sigma^2_O$, which indicates the variance of constant errors associated with observers (e.g., some observers are more lenient than others). The residual component has mean zero and variance $\sigma^2_{KO,e}$, which indicates the degree to which observers score particular units differently along with residual error due to unidentified facets or randomness. The collection of observed scores, **X**, has a variance $\sigma^2_X = \sum(X_{ko} - \mu)^2$, which equals the sum of the variance components:

$$\sigma^2_X = \sigma^2_K + \sigma^2_O + \sigma^2_{KO,e} \qquad \text{(A.2)}$$

G theory focuses on these variance components. The relative magnitudes of the components provide information about particular sources of error influencing a measurement. It is convenient to estimate variance components from the analysis of variance (ANOVA) using sample data (unit-performance scores). Numerical estimates of the variance

components are obtained by setting the expected mean squares equal to the observed mean squares and solving the set of simultaneous equations as shown in Table A.1.

Table A.1

ESTIMATES OF VARIANCE COMPONENTS FOR A ONE-FACET, $K \times O$ DESIGN

Source of Variation	Mean Square	Expected Mean Square[a]	Estimated Variance Component
Unit (K)	MS_K	$\sigma^2_{KO,e} + n_O\sigma^2_K$	$\sigma^2_K = (MS_K - MS_{res})/n_O$
Observer (O)	MS_O	$\sigma^2_{KO,e} + n_K\sigma^2_K$	$\sigma^2_O = (MS_O - MS_{res})/n_K$
$K \times O, e$	MS_{res}	$\sigma^2_{KO,e}$	$\sigma^2_{KO,e} = MS_{res}$

[a] n_O = number of observers; n_K = number of units.

G theory distinguishes a *decision* (D) study from a *generalizability* (G) study. This distinction recognizes that certain studies are associated with the development of a measurement procedure (G studies) whereas other studies then apply the procedure (D studies). In planning the D study, the decisionmaker (i) defines the universe of generalization and (ii) specifies his proposed interpretation of a measurement. These plans determine (iii) the questions to be asked of the G study data in order to optimize the measurement design. Each of these points is considered in turn.

(i) G theory recognizes that the universe of admissible observations encompassed by a G study may be broader than the universe to which a decisionmaker wishes to generalize. That is, the decisionmaker proposes to generalize to a universe comprised of some subset of facets in the G study. This universe is called the *universe of generalization*. It may be defined by reducing the universe of admissible observations, i.e., by reducing the levels of a facet (e.g., creating a fixed facet, called a fixed factor in ANOVA), by selecting one level of a facet, and thereby controlling it, or by ignoring a facet. All three alternatives have consequences for the estimation of the components of error variance that enter into the observed score variance.

(ii) G theory recognizes that decisionmakers use the same unit performance measurement in different ways. For example, some interpretations may focus on differences between units (i.e., relative or comparative decisions), some may use the observed score as an estimate of

a unit's universe score (absolute decisions), while still others may use the observed score in a regression estimate of the universe score (see, for example, Kelley's (1947) regression estimate of true scores). There is a different error associated with each of these proposed interpretations. For *relative decisions*, the error in a $K \times O$ design is defined as

$$\delta_k = (X_{k\cdot} - \mu_{\cdot}) - (\mu_k - \mu) \tag{A.3}$$

where the dot indicates that an average has been taken over the levels of the observer facet (o) under which the unit (k) was observed. The variance of the errors for relative decisions is

$$\sigma_\delta^2 = \sigma_{k\cdot}^2 = \sigma_{ko,e}^2 / n_O' \tag{A.4}$$

where n_O' indicates the number of conditions of facet o to be sampled in a D study. Notice that (a) $\sigma_{ko,e}^2 / n_O'$ is the standard error of the mean of a unit's scores averaged over the observers, and (b) the magnitude of the error is under the control of the decisionmaker. To reduce σ_δ^2, n_O' may be increased.

For *absolute decisions*, the error is defined as

$$\Delta_k = X_{k\cdot} - \mu_k \tag{A.5}$$

The variance of these errors in a $K \times O$ design is

$$\sigma_\Delta^2 = \sigma^2 + \sigma_{k\cdot}^2 = \sigma_O^2 / n_O' + \sigma_{ko,e}^2 / n_O' \tag{A.6}$$

In contrast to σ_δ^2, σ_Δ^2 includes the variance of constant errors associated with facet o (σ_O^2). This arises because, in absolute decisions, the leniency of the particular observer that a unit receives will influence the unit's observed score and, hence, the decisionmaker's estimate of the unit's universe score. For relative decisions, however, the effect of observer is constant for all units and so does not influence the rank ordering of them (see Erlich and Shavelson, 1976).

Finally, for decisions based on the *regression estimate* of a unit's universe score, error (of estimate) is defined as

$$\epsilon_k = \hat{\mu}_k - \mu_k \tag{A.7}$$

where $\hat{\mu}_k$ is the regression estimate of a unit's universe score, μ_k. The estimation procedures for the variance of errors of estimate may be found in Cronbach et al. (1972, pp. 97ff).

(iii) D studies encompass a wide variety of designs including crossed, partially nested, and completely nested designs. All facets in the D design may be random (random model), or only some may be random (mixed model). Often, in D studies, nested designs are used for convenience, for increasing sample size, or both. Observers may be nested within units, i.e., one team of observers evaluates the performance of unit 1, a second team unit 2, and so on (we write $o(k)$ or $o{:}k$ to denote nesting). So, the effect of the constant errors associated with facet o is confounded with the effect associated with the unit by o-facet interaction (ko,e).

$$\sigma^2_{X_{k.}} = \sigma^2_k + \sigma^2_{k.e} = \sigma^2_k + \sigma^2_\Delta \tag{A.8}$$

Note that, for a completely nested design, $\sigma^2_\delta = \sigma^2_\Delta$.

While stressing the importance of variance components and errors such as σ^2_δ, G theory also provides a coefficient analogous to the reliability coefficient in classical theory. The *generalizability coefficient*, ρ^2, is defined as the ratio of the universe-score variance to the expected observed-score variance, i.e., an intraclass correlation:

$$\rho^2 = \sigma^2_k / \sigma^2(X) = \sigma^2_k / (\sigma^2_k + \sigma^2_\delta) \tag{A.9}$$

The expected observed-score variance is used in G theory because the theory assumes only random sampling of the levels of facets and so the observed-score variance may change from one application of the design to another. Sample estimates of the parameters in Eq. (A.9) are used to estimate the G coefficient:

$$\hat{\rho}^2 = \hat{\sigma}^2_k / (\hat{\sigma}^2_k + \hat{\sigma}^2_\delta) \tag{A.9a}$$

$\hat{\rho}^2$ is a biased but consistent estimator of ρ^2.

For absolute decisions a generalizability coefficient can be defined in an analogous manner:

$$\rho^2 = \sigma^2_k / (\sigma^2_k + \sigma^2_\Delta) \tag{A.10}$$

$$\hat{\rho}^2 = \hat{\sigma}^2_k / (\hat{\sigma}^2_k + \hat{\sigma}^2_\Delta) \tag{A.10a}$$

Finally, note that, for completely nested designs regardless of whether relative or absolute decisions are to be made, error variance is defined as σ^2_Δ, and so Eq. (A.10) provides the generalizability coefficient for such designs.

Description of the Tank Crew Data

The data on tank crew performance were collected in the spring of 1971 during the annual qualification firing exercises at the Seventh Army Training Center, Grafenwohr, Germany. The tank crews performed the Table VIII mission—Deliberate Attack (Live Fire)—based on the Army Training and Evaluation Program for Mechanized Infantry Tank Force (ARTEP 71-2, 1978). The tank crews represented three companies, with each company comprised of three platoons of five tank crews each. The performance of the 45 tank crews was scored on two occasions, once when they carried out the mission in daytime and once at night. A single observer scored the performance of each crew according to the detailed ARTEP guidelines. Scores for the sample of 45 crews ranged from a low of 210 to a high of 1150. We refer to these scores hereafter as Table VIII data.

In using this data set to demonstrate the applicability of G theory to assessing the reliability (generalizability) of unit performance measurements, a caveat is in order. Although performance was measured on two occasions, those occasions differed over days, time of day, (probably) in the observer who assigned the performance score, and in other unknown ways. The use of such confounded data is contrary to generalizability theory. Indeed, generalizability theory stresses that each of these facets—time, time of day, observer, etc.—should be measured and their effect on the reliability of the data estimated. We use these data, then, heuristically. They represent the hierarchical nature of unit performance data and provide a numerical example for demonstrating G theory's applicability.

Classical Reliability of Unit Performance Data

In classical theory, the best estimate we have of the reliability of the tank crew performance data is the correlation between tank crew scores obtained on two occasions, once during the day and once at night. For a performance score averaged over the two occasions and ignoring the effect of platoon and company, the reliability is 0.639. (The pooled, within-company reliability is 0.680.)

Clearly, this reliability coefficient is influenced by the leniency of different observers, the difficulty of the terrain or terrains on which the missions were conducted, the differences between missions, the time (day or night), the day that the performance was observed, and so forth. However, we have no way to estimate the importance of these possible sources of measurement error using classical reliability theory even if the facets of these measurements had been systematically iden-

tified and manipulated. Furthermore, performance might be influenced by the policies and leadership skills within particular companies or platoons. Classical reliability theory is mute on how to treat these hierarchical data.

Dependability of Unit Performance Measurements

G theory provides a means of treating the typical, hierarchically nested units of analysis found in military data. In the Table VIII data, crews are nested in platoons which are nested in companies. G theory first requires identification of the decisionmaker or at least the level in the hierarchy (crew, platoons, or companies) on which decisions will bear. Some decisionmakers, for example, may be interested in crew performance, whereas others may be interested in the performance of platoons or companies. The point of interest has implications for estimating the generalizability (reliability) of unit performance measurements. If the decisionmaker is interested in crew performance and wishes to generalize, say, to the performance of those crews over missions, observers, and days, then the generalizability of the performance measurements will refer to the systematic variation between tank crews due to nonchance differences between the crews themselves, between the platoons in which they operate, and between the companies in which the platoons operate (Cardinet, Toureur, and Allal, 1981; for a review, see Shavelson and Webb, 1981). Now, consider the case where the decisionmaker is interested in the performance of platoons. In this case, the set of platoons nested within companies forms the population of interest, and systematic variability in the performance of tank crews nested within the platoons introduces error into the measurement of platoon performance.

To demonstrate the application of G theory to hierarchical populations, we use the Table VIII data set with crews (k=5) nested within platoons, and platoons (j=3) nested within each company (i=3). Performance is measured on two different occasions (l=2 occasions, day and night). In shorthand form, we write this design as companies × platoons (companies) × crews (platoons:companies) × occasions.

For each source of variation in this measurement design, the underlying variance components can be determined by taking the expectation of the mean squares (see Table A.2). By setting the expected mean squares equations equal to their corresponding observed mean squares, we can solve for each variance component, as has been done in Table A.3 using Table VIII data. These variance components provide estimates of the magnitude of error contributed by each facet of meas-

Table A.2

EXPECTED MEAN SQUARES IN A $C \times P(C) \times Cr(PC) \times O$ GENERALIZABILITY DESIGN

Source	Expected Mean Squares
Companies *(C)*	$\sigma^2_{res} + k\sigma^2_{PO} + jk\sigma^2_{CO} + l\sigma^2_{Cr} + kl\sigma^2_{P} + jkl\sigma^2_{C}$
Platoons *(P:C)*	$\sigma^2_{res} + k\sigma^2_{PO} + l\sigma^2_{Cr} + kl\sigma^2_{P}$
Crews *(Cr:P:C)*	$\sigma^2_{res} + l\sigma^2_{Cr}$
Observers *(O)*	$\sigma^2_{res} + k\sigma^2_{PO} + jk\sigma^2_{CO} + ijk\sigma^2_{O}$
$C \times O$	$\sigma^2_{res} + k\sigma^2_{PO} + jk\sigma^2_{CO}$
$P{:}C \times O$	$\sigma^2_{res} + k\sigma^2_{PO}$
$Cr{:}P{:}C \times O, e$	σ^2_{res}

Table A.3

VARIANCE COMPONENTS FROM THE TABLE VIII DATASET

Source	Mean Square	Estimated Variance Component
Companies *(C)*	55461	0[a]
Platoons *(P:C)*	78636	1607.19
Crews *(Cr:P:C)*	45383	15967.50
Occasions *(O)*	244505	3573.21
$C \times O$	83711	3538.79
$P{:}C \times O$	30629	3436.17
$Cr{:}P{:}C \times O$ *(res)*	31448	13448.20

[a]Negative variance component set to zero.

urement as well as an estimate of the systematic variation due to the focus of measurement (e.g., tank crews).

In theory, a variance component cannot be negative. With sample data, such as the Table VIII data, a negative variance component can arise either due to sampling error or misspecification of the measurement model. If the former, the most widely accepted practice is to set

the variance component to zero, as we have done in Table A.3. If the latter, the model should be respecified and variance components estimated with the new model. Our rationale for setting the company variance component to zero is the following. First, the difference in mean performance of the three companies is small: 770.90, 763.33, and 692.93. Variation among company means accounts for only 0.3 percent of the total variation in the data. We believe, then, that the best estimate of the variance due to companies is zero and the difference among sample means represents sampling error. (For a review of the substantive and statistical issues regarding negative variance components, see Shavelson and Webb, 1981.)

The largest variance component in Table A.3 is associated with crews; crew performance differs systematically, and the measurement procedure reflects this systematic difference. The next largest component is associated with the residual, indicating that error is introduced due to inconsistency in tank crew performance from one occasion to the next and other unidentified sources of error (e.g., inconsistency due to time of day, observer, terrain, and the like). The remaining variance components are roughly one-fourth the size of the residual, with the exception of the component for companies.

Since the variance component for companies is 0 and the variance component for platoons is the smallest one remaining, we conclude that neither sufficiently influences variation among crews enough to have an important influence on the systematic differences between crews.

Since decisionmakers are interested in the reliability of unit performance, one possible method for calculating the generalizability coefficient for crews is

$$\rho^2(crews) = \frac{\sigma^2_{Cr(P:C)}}{\sigma^2_{Cr(P:C)} + \frac{\sigma^2_O}{n_O} + \frac{\sigma^2_{res}}{n_O}} = 0.65 \qquad \text{(A.11)}$$

The generalizability of a tank crew's performance, averaged over the two observation occasions (day and night), is 0.65. If, however, the decisionmaker is interested in the generalizability of the score of a single tank crew selected randomly and observed on one occasion, the reliability drops to 0.48. This large drop in generalizability (reliability) is due to the large residual of $(C) \times (P : C) \times (Cr : P : C) \times$ occasions and other unidentified sources of error.

Cardinet et al. (1981) argue, and we concur, that the universe-score variance is comprised of all components that give rise to systematic variation between crews. In this case, variation due to companies and

platoons, as well as variation due to crews, would be considered universe-score variation. Characteristics of companies and platoons, such as leadership ability, contribute to systematic variation between crews. Following Cardinet et al., this generalizability coefficient for crews, averaged over two observation occasions, is

$\rho^2(crews^*)$ =

$$\frac{\sigma^2_{Cr(P:C)} + \sigma^2_{P(C)} + \sigma^2_{C}}{\sigma^2_{Cr(P:C)} + \sigma^2_{P(C)} + \sigma^2_{C} + \frac{\sigma^2_{O}}{n_O} + \frac{\sigma^2_{P(C)O}}{n_O} + \frac{\sigma^2_{CO}}{n_O} + \frac{\sigma^2_{res}}{n_O}} = 0.59 \quad \text{(A.11a)}$$

We write $\rho^2(crews^*)$ to distinguish this coefficient from the one in Eq. (A.11).

Surprisingly, by increasing the universe-score variances (i.e., Eq. (A.11a)), the generalizability coefficient decreased, for two reasons. The increase in universe-score variance by incorporating systematic variation due to companies and platoons was negligible

$$\sigma^2_{C} = 0,\ \sigma^2_{P(C)} = 1607.19.$$

And the additional error introduced ($\sigma^2_{P(C)O}$ and σ^2_{CO}) by incorporating these sources of universe-score variance, while not large relative to other sources of error variance (e.g., σ^2_{res}), were large relative to the systematic variability of companies and platoons.

If the decisionmaker is interested in platoon performance, the generalizability of the measurement can be estimated (aggregating over crews within platoons and occasions) as follows:

$\sigma^2(platoons)$ =

$$\frac{\sigma^2_{P(C)}}{\sigma^2_{P(C)} + \frac{\sigma^2_{Cr(P:C)}}{n_{Cr(P:C)}} + \frac{\sigma^2_{O}}{n_O} + \frac{\sigma^2_{P(C)O}}{n_O} + \frac{\sigma^2_{res}}{n_{Cr(P:C)}n_O}} = 0.17 \quad \text{(A.12)}$$

Notice here that *crews* is considered a source of error—variability in crews introduces error in estimating the performance of the entire platoon, the average (or sum) of the performance of a platoon's individual crews. Indeed, variation among crews constitutes a major source of error. The low generalizability coefficient, then, reflects the fact that there is greater variability among crews within a platoon than there is variability among platoons.

Validity refers to the accuracy of a proposed interpretation of a measurement. For example, the following question might be raised of a reliable (generalizable) unit performance measurement: "How accurate is this measurement for deciding whether a mechanized infantry unit is likely to accomplish its mission in wartime?" Whereas a generalizability coefficient tells you how dependable a measurement is from one occasion to the next, one observer to the next, etc., a validity coefficient provides an index of the accuracy of the proposed interpretation of that measurement. One way to examine the validity of, for example, Table VIII scores would be to observe tank crews in the Table VIII simulation and the same crews performing a mission during wartime. While this is unlikely, it points out that, ultimately, we use Table VIII scores to tell us something about how well tank crews would perform in a wide variety of conditions and missions that are important.

Criterion situations such as wartime, of course, are not available for validation purposes. Compromises, then, are made in validity studies. One such compromise is to observe tank crew performance in a simulated wartime setting. Tank crews might be observed under a wide variety of missions against a wide variety of different opposing units in a "war game." Although this is not the criterion situation, war, it may provide a fairly high fidelity simulation of one small part of a wartime operation.

A second method for validating an interpretation is to measure unit performance by different methods. This is akin to the astronomer's method of triangulation. If very different methods of measurement agree, confidence is increased in the interpretation of the performance measurement. For example, the data set described above contained Table VIII scores and ratings of tank crew performance by an expert with many years of mechanized infantry battlefield experience. Here are two somewhat different measurement methods. Table VIII scores are based on observers' records of, for example, the number of enemy casualties; the scale has a sample mean of 192.45 and a standard deviation of 115.97. The expert rated overall performance on a three-point scale: below criterion (0), above criterion (1), and excellent (2) (for this analysis, we collapsed the last two categories and so the scale takes on two values, 0 and 1, $\bar{x} = 0.5$ and s.d. = 0.5). Even though both measures are less than "maximally different" as theory dictates, they are getting at the same thing, the performance of tank crews (perhaps) in wartime. The correlation between the two measures of tank crew performance, then, should be high. It was (0.83); this increases our confidence that the Table VIII measures captured some important aspects of tank crew combat performance.

More generally, validation is a process of amassing evidence that increases our confidence in the accuracy of the proposed interpretation. This process is to pose counterinterpretations to the proposed interpretation and collect data to see if the proposed interpretation holds up. In this respect, we cannot set forth all possible methods for validating unit performance measurements. Validation depends on the proposed interpretation and counterinterpretations.

REFERENCES

Adams, J., et al., "Group Performance at West Point," *Armed Forces and Society*, Vol. 7, 1981, pp. 246–255.

Adams, S., "Status Congruency as a Variable in Small Group Performance," *Social Forces*, Vol. 32, 1953, pp. 16–22.

Allen, R. F., and S. Silverzweig, "Group Norms: Their Influence on Training Effectiveness," in R. L. Craig (ed.), *Training and Development Handbook*, McGraw-Hill Book Co., New York, 1976.

Altman, I., and W. W. Haythorn, "The Effects of Social Isolation and Group Composition on Performance," *Human Relations*, Vol. 20, 1967, pp. 313–340.

Bakeman, R., and R. Helmreich, "Cohesiveness and Performance: Covariation and Causality in an Undersea Environment," *Journal of Experimental Social Psychology*, Vol. 11, 1975, pp. 478–489.

Ball, J. R., and A. V. Carron, "The Influence of Team Cohesion and Participation Motivation Upon Performance Success in Intercollegiate Ice Hockey," *Canadian Journal of Applied Sport Sciences*, Vol. 1, 1976, pp. 271–275.

Bass, B. M., "Team Productivity and Individual Member Competence," *Small Group Behavior*, Vol. 11, 1980, pp. 431–504.

Bauer, R. G., R. L. Stout, and R. F. Holz, *Developing a Conceptual and Predictive Model of Discipline in the U.S. Army*, ARI Research Problem Review 76-5, Alexandria, Virginia, September 1976.

Berkowitz, L., "Group Norms Among Bomber Crews: Patterns of Perceived Crew Attitudes, 'Actual' Crew Attitudes, and Crew Liking Related to Aircrew Effectiveness in Far Eastern Combat," *Sociometry*, Vol. 19, 1956, pp. 141–153.

Berkowitz, L., and B. I. Levy, "Pride in Group Performance and Group-Task Motivation," *Journal of Abnormal and Social Psychology*, Vol. 53, 1956, pp. 300–306.

Bessinger, G. J., Jr., *Correlation of Basic Combat Trainee Morale and Attitude with Performance Data*, U.S. Army Training Center, Directorate of Organization Development, Fort Ord, California, July 1974.

Bird, A. M., "A Group Dynamics Approach to Effective Coaching of Team Sports," *Motor Skills: Theory into Practice*, Vol. 2, 1978, pp. 92–101.

Bird, A. M., "Development of a Model for Predicting Team Performance," *Research Quarterly*, Vol. 48, 1977, pp. 24–32.

Bird, A. M., and J. M. Brame, "Self Versus Team Attributions: A Test of the 'I'm OK, but the Team's So-So' Phenomenon," *Research Quarterly*, Vol. 49, 1978, pp. 260–268.

Bird, A. M., C. D. Foster, and G. Maruyama, "Convergent and Incremental Effects of Cohesion on Attributions for Self and Team," *Journal of Sport Psychology*, Vol. 2, 1980, pp. 181–194.

Black, B., *ASVAB Aptitude Area Score, CO, as a Predictor of Tank Crewmember Performance*, ARI Fort Knox Field Unit FKFU Working Paper 80-9, Fort Knox, Kentucky, October 1980.

Bons, P. M., and F. E. Fiedler, "Changes in Organizational Leadership and the Behavior of Relationship- and Task-Motivated Leaders," *Administrative Science Quarterly*, Vol. 21, 1976, pp. 453–473.

Bouchard, T. J., Jr., "Personality, Problem-Solving Procedure, and Performance in Small Groups," *Journal of Applied Psychology Monograph*, Vol. 53, No. 1, Pt. 2, 1969.

Bouchard, T. J., Jr., "Training, Motivation, and Personality as Determinants of the Effectiveness of Brainstorming Groups and Individuals," *Journal of Applied Psychology*, Vol. 56, 1972, pp. 324–331.

Bowen, D. D., and J. P. Siegel, "Process and Performance: A Longitudinal Study of the Reactions of Small Task Groups to Periodic Performance Feedback," *Human Relations*, Vol. 26, 1973, pp. 433–448.

Burstein, K. R., "What Is Learned: All Kinds of Things," *Bulletin of the Psychonomic Society*, Vol. 16, 1980, pp. 232–234.

Butler, M. C., and Ralph G. Burr, "Utility of a Multidimensional Locus of Control Scale in Predicting Health and Job-Related Outcomes in Military Environments," *Psychological Reports*, Vol. 47, 1980, pp. 719–728.

Byrne, D., "Can Wright Be Wrong? Let Me Count the Ways," *Representative Research in Social Psychology*, Vol. 2, 1971, pp. 12–18.

Cardinet, J., Y. Toureur, and L. Allal, "Extension of Generalizability Theory and Its Applications in Educational Measurement," *Journal of Educational Measurement*, Vol. 18, 1981, pp. 183–204.

Carron, A. V., *Social Psychology of Sport*, Mouvement Publications, Ithaca, New York, 1980.

Comrey, A. L., "Group Performance in a Manual Dexterity Task," *Journal of Applied Psychology*, Vol. 37, 1953, pp. 207–210.

Comrey, A. L., and C. K. Staats, "Group Performance in a Cognitive Task," *Journal of Applied Psychology*, Vol. 39, 1955, pp. 354–356.

Cook, D. M., "The Impact of Managers of Frequency of Feedback," *Academy of Management Journal*, Vol. 11, No. 3, September 1968, pp. 263–277.

Cooper, C. L., *Theories of Group Processes*, John Wiley and Sons, Inc., New York, 1975.

Cooper, R., and R. Payne, "Personality Orientations and Performance in Soccer Teams," *British Journal of Social and Clinical Psychology*, Vol. 11, 1972, pp. 2–9.

Cory, B., et al., *Duty Modules: An Approach to the Identification and Classification of Personnel Resources and Requirements*, ARI Technical Paper 367, Alexandria, Virginia, June 1979.

Cronbach, L. J., G. C. Gleser, H. Nanda, and N. Rajaratnam, *The Dependability of Behavioral Measurements*, John Wiley and Sons, Inc., New York, 1972.

Dean, L. M., et al., "Manning Levels, Organizational Effectiveness, and Health," *Human Relations*, Vol. 32, 1979, pp. 237–246.

Department of the Army, *Army Training and Evaluation Program for Mechanized Infantry Tank Task Force*, ARTEP 71-2, U.S. Army Adjutant General Publications Center, Baltimore, Maryland, September 1978.

Dyer, J. L., T. R. Tremble, Jr., and D. L. Finley, *The Structural, Training, and Operational Characteristics of Army Teams*, ARI Technical Report 507, Alexandria, Virginia, June 1980.

Eaton, N. K., *Predicting Tank Gunnery Performance*, ARI Research Memorandum 78-6, Alexandria, Virginia, February 1978a.

Eaton, N. K., *Tank Crew Stability and Tank Gunnery Performance*, ARI Research Memorandum 78-15, Alexandria, Virginia, July 1978b.

Eaton, N. K., *Performance Motivation in Armor Training*, ARI Technical Paper 291, Alexandria, Virginia, September 1978c.

Eaton, N. K., D. W. Bessemer, and D. M. Kristiansen, *Tank Crew Position Assignment*, ARI Technical Report 391, Alexandria, Virginia, October 1979.

Eaton, N. K., and J. F. Neff, *The Effects of Tank Crew Turbulence on Tank Gunnery Performance*, ARI Technical Paper 350, Alexandria, Virginia, September 1978.

Eaton, N. K., J. R. Johnson, and B. A. Black, *Job Samples as Tank Gunnery Performance Predictors*, ARI Technical Report 473, Alexandria, Virginia, September 1980.

Egerman, K., "Effects of Team Arrangement on Team Performance," *Journal of Personality and Social Psychology*, Vol. 3, 1966, pp. 541–550.

Eitzen, D. S., "Group Structure and Group Performance," in W. F. Straub (ed.), *Sport Psychology: An Analysis of Athlete Behavior*, Mouvement Publications, Ithaca, New York, 1978.

Eitzen, D. S., "The Effect of Group Structure on the Success of Athletic Teams," *International Review of Sport Sociology*, Vol. 8, 1973, pp. 7–17.

Ellsworth, R. B., "Feedback: Asset or Liability in Improving Treatment Effectiveness?" *Journal of Consulting and Clinical Psychology*, Vol. 40, 1973, pp. 383–393.

Erlich, O., and R. J. Shavelson, *The Applicability of Generalizability Theory to Research on Teaching*, Technical Report 76-9-1, Beginning Teacher Evaluation Study, Far West Laboratory, September 1976.

Feldman, R. A., "Group Integration and Intense Interpersonal Disliking," *Human Relations*, Vol. 22, 1969, pp. 405–413.

Fiedler, F. E., "The Psychological-Distance Dimension in Interpersonal Relations," *Journal of Personality*, Vol. 22, 1953, pp. 142–150.

Fiedler, F. E., "Assumed Similarity Measures as Predictors of Team Effectiveness," *Journal of Abnormal and Social Psychology*, Vol. 49, 1954, pp. 381–388.

Fiedler, F. E., "Contingency Model of Leadership Effectiveness," in L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, Vol. I, Academic Press, New York, 1964, pp. 149–190.

Fiedler, F. E., *A Theory of Leadership Effectiveness*, McGraw-Hill Book Co., New York, 1967.

Fiedler, F. E., "The Contingency Model and the Dynamics of the Leadership Process," in L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, Vol. 11, Academic Press, New York, 1978, pp. 60–112.

Fiedler, F. E., and A. F. Leister, "Leader Intelligence and Task Performance: A Test of a Multiple Screen Model," *Organizational Behavior and Human Performance*, Vol. 20, 1977, pp. 1–14.

Fiedler, F. E., et al., "Organizational Stress and the Use and Misuse of Managerial Intelligence and Experience," *Journal of Applied Psychology*, Vol. 64, 1979, pp. 635–647.

Foeldesi, T., "Study on Favoured Sociological and Sociopsychological Factors Influencing Performance of Athletic Teams in Rowing," *International Review of Sport Sociology*, Vol. 11, 1976, pp. 17–32.

Gavin, J. F., and J. G. Howe, "Psychological Climate: Some Theoretical and Empirical Considerations," *Behavioral Science*, Vol. 20, 1975, pp. 228–240.

Ghiselli, E. E., "The Forced-Choice Technique in Self Description," *Personnel Psychology*, Vol. 7, 1954, pp. 201–208.

Ghiselli, E. E., and T. Lodahl, "Patterns of Managerial Traits and Group Effectiveness," *Journal of Abnormal and Social Psychology*, Vol. 57, 1958, pp. 61–66.

Gill, D. L., "Prediction of Group Motor Performance from Individual Member Abilities," *Journal of Motor Behavior*, Vol. 11, 1979, pp. 113–122.

Goldin, S. E., and P. W. Thorndyke, "An Overview of Workshop Issues," in S. E. Goldin and P. W. Thorndyke (eds.), *Improving Team Performance: Proceedings of The Rand Team Performance Workshop*, The Rand Corporation, R-2606-ONR, August 1980.

Goldman, M., "A Comparison of Individual and Group Performance for Varying Combinations of Initial Ability," *Journal of Personality and Social Psychology*, Vol. 1, No. 3, March 1965, pp. 210–216.

Goldman, M., and L. A. Fraas, "The Effects of Leader Selection on Group Performance," *Sociometry*, Vol. 28, 1965, pp. 82–88.

Goodacre, D. M., III, "Group Characteristics of Good and Poor Performing Combat Units," *Sociometry*, Vol. 16, 1953, pp. 168–179.

Goodacre, D. M., III, "The Use of a Sociometric Test as a Predictor of Combat Unit Effectiveness," *Sociometry*, Vol. 14, 1951, pp. 148–152.

Gottheil, E., and D. P. Vielhaber, "Interaction of Leader and Squad Attributes Related to Performance of Military Squads," *Journal of Social Psychology*, Vol. 68, 1966, pp. 113–127.

Gruber, J. J., and G. R. Gray, "Factor Patterns of Variables Influencing Cohesiveness at Various Levels of Basketball Competition," *Research Quarterly for Exercise and Sport*, Vol. 52, 1981, pp. 19–30.

Grunzke, P. M., *Evaluation of the Automated Adaptive Flight Training System's Air-to-Air Intercept Performance Measurement*, Air Force Human Resource Laboratories, AFHRL-TR-76-23, Williams AFB, Arizona, July 1978.

Hagan, C., "Improved ARTEP in Reforger 1981 and the Future," in R. H. Sulzen (ed.), *Unit Performance Measurement: Proceedings of the ARI Sponsored Seminar*, ARI, Alexandria, Virginia, October 1981.

Hagstrom, W. O., and H. C. Selvin, "Two Dimensions of Cohesiveness in Small Groups," *Sociometry*, Vol. 28, 1965, pp. 30–43.

Hall, E. R., and W. A. Rizzo, *An Assessment of U.S. Navy Tactical Team Training*, Training Analysis and Evaluation Group Report No. 18, Orlando, Florida, March 1975.

Hare, A. P., *Handbook of Small Group Research*, The Free Press, New York, 1976.

Havron, M. D., L. H. McFarling, and R. G. Wanschura, *Improved Army Training and Evaluation (ARTEP) Methods for Unit Evaluation: Volume VI: Conventional ARTEP Methods and Engagement Simulations: An Examination of Options*, ARI Technical Report 9-A24, Alexandria, Virginia, April 1979.

Hawley, K. E., and J. S. Heinen, "Compatibility and Task Group Performance," *Human Relations*, Vol. 32, 1979, pp. 579–590.

Haythorn, W., "The Influence of Individual Members on the Characteristics of Small Groups," *Journal of Abnormal and Social Psychology*, Vol. 48, 1953, pp. 276–284.

Heslin, R., "Predicting Group Task Effectiveness from Member Characteristics," *Psychological Bulletin*, Vol. 62, 1964, pp. 248–256.

Hewett, T. T., G. E. O'Brien, and J. Hornik, "The Effects of Work Organization, Leadership Style, and Member Compatibility upon the Productivity of Small Groups Working on a Manipulative Task," *Organizational Behavior and Human Performance*, Vol. 11, 1974, pp. 283–301.

Hill, R. E., "Interpersonal Compatibility and Workgroup Performance," *The Journal of Applied Behavioral Science*, Vol. 11, 1975, pp. 210–219.

Hoffman, L. R., "Homogeneity of Member Personality and Its Effect on Group Problem-Solving," *Journal of Abnormal and Social Psychology*, Vol. 58, 1959, pp. 27–32.

Hoffman, L. R., and N.R.F. Maier, "Quality and Acceptance of Problem Solutions by Members of Homogeneous and Heterogeneous Groups," *Journal of Abnormal and Social Psychology*, Vol. 62, 1961, pp. 401–407.

Hoffman, L. R., and C. G. Smith, "Some Factors Affecting the Behavior of Members of Problem Solving Groups," *Sociometry*, Vol. 23, 1960, pp. 273–291.

Horsfall, A. B., and C. M. Arensberg, "Teamwork and Productivity in a Shoe Factory," *Human Organization*, Vol. 8, 1949, pp. 13–25.

Iso-Ahola, S., "Determinants of Evaluation of Team Performance," *Scandinavian Journal of Psychology*, Vol. 17, 1976, pp. 292–296.

Janis, I. L., *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*, 2d ed., Houghton-Mifflin, Boston, 1983.

Jones, A. P., and L. R. James, "Psychological Climate: Dimensions and Relationships of Individual and Aggregated Work Environment Perceptions," *Organizational Behavior and Human Performance*, Vol. 23, 1979, pp. 201–250.

Jones, M. B., "Regressing Group on Individual Effectiveness," *Organizational Behavior and Human Performance*, Vol. 11, 1974, pp. 426–451.

Kabanoff, B., and G. E. O'Brien, "Cooperation Structure and the Relationship of Leader and Member Ability to Group Performance," *Journal of Applied Psychology*, Vol. 64, 1979, pp. 526–532.

Kahan, J. P., and J. P. Poitou, "Single Linear Ordering Versus Multiple Discrepant Orderings in the Cognitive Representation of Social Structures," *European Journal of Social Psychology*, Vol. 3, 1973, pp. 109–124.

Kelley, T. L., *Fundamentals of Statistics*, Harvard University Press, Cambridge, Massachusetts, 1947.

Kim, J. S., and W. C. Hamner, "Effect of Performance Feedback and Goal Setting on Productivity and Satisfaction in an Organizational Setting," *Journal of Applied Psychology*, Vol. 61, No. 1, February 1976, pp. 48–57.

Klaus, D. J., and R. Glaser, *Increasing Team Proficiency through Training: 5. Team Learning as a Function of Member Learning Characteristics and Practice Conditions*, Report AIR-E1-4/65-TR, American Institute for Research, Pittsburgh, May 1965.

Klein, M., and G. Christiansen, "Gruppenkomposition, Gruppenstruktur, und Effektivitaet von Basketball (Group Composition, Group Structure, and Effectiveness in Basketball)," *Koelner Zeitschrift fuer Soziologie und Sozialpsychologie*, Vol. 10, 1966, pp. 180–191.

Klemp, G. O., Jr., M. T. Munger, and L. M. Spencer, Jr., *Analysis of Leadership and Management Competencies of Commissioned and Noncommissioned Naval Officers in the Pacific and Atlantic Fleets*, McBer and Company, Task Planning Document EG-33, Boston, Massachusetts, August 1977.

Knerr, C. M., R. T. Root, and L. E. Word, *An Application of Tactical Engagement Simulation for Unit Proficiency Measurement*, ARI Technical Paper 381, Alexandria, Virginia, July 1979.

Lampkin, E. C., "Effects of n-dominance and Group Composition on Task Efficiency in Laboratory Triads," *Organizational Behavior and Human Performance*, Vol. 7, 1972, pp. 189–202.

Landers, D. M., and T. F. Crum, "The Effect of Team Success and Formal Structure on Interpersonal Relations and Cohesiveness of Baseball Teams," *International Journal of Sport Psychology*, Vol. 2, 1971, pp. 88–96.

Landers, D. M., and G. Lueschen, "Team Performance Outcome and the Cohesiveness of Competitive Coacting Groups," *International Review of Sport Sociology*, Vol. 9, 1974, pp. 57–71.

Landers, D. M., L. R. Brawley, and D. M. Landers, "Group Performance, Interaction, and Leadership," in G.R.F. Lueschen and G. H. Sage (eds.), *Handbook of Social Science of Sport*, Stipes, Champaign, Illinois, 1981, pp. 297–315.

Laughlin, P. R., and L. G. Branch, "Individual vs. Tetradic Performance on a Complementary Task as a Function of Initial Ability Level," *Organizational Behavior and Human Performance*, Vol. 8, 1972, pp. 201–216.

Laughlin, P. R., L. G. Branch, and H. H. Johnson, "Individual vs. Triadic Performance on a Unidimensional Complementary Task as a Function of Initial Ability Level," *Journal of Personality and Social Psychology*, Vol. 12, No. 2, June 1969, pp. 144–150.

Liddell, W. W., and J. W. Slocum, "The Effects of Individual-Role Compatibility upon Group Performance: An Extension of Schutz's FIRO Theory," *Academy of Management Journal*, Vol. 19, No. 3, September 1976, pp. 413–426.

Lodahl, T., and L. W. Porter, "Psychometric Score Patterns, Social Characteristics, and Productivity in Small Industrial Work Groups," *Journal of Applied Psychology*, Vol. 45, No. 2, April 1961, pp. 73–79.

Lord, R. G., and M. Rowzee, "Task Interdependence, Temporal Phase, and Cognitive Heterogeneity as Determinants of Leadership-Behavior and Behavior-Performance Relations," *Organizational Behavior and Human Performance*, Vol. 23, 1979, pp. 182–200.

Lott, A. J., and B. E. Lott, "Group Cohesiveness as Interpersonal Attraction: A Review of Relationships with Antecedent and Consequent Variables," *Psychological Bulletin*, Vol. 64, 1965, pp. 259–309.

MacCrimmon, K. R., "Studying Team Performance: A Decision Theory Perspective," in S. E. Goldin and P. W. Thorndyke (eds.), *Improving Team Performance: Proceedings of The Rand Team Performance Workshop*, The Rand Corporation, R-2606-ONR, August 1980.

Madden, J. L., "Tactical Training Development Systems and Systems Integration Centers," in R. H. Sulzen (ed.), *Unit Performance Measurement: Proceedings of the ARI Sponsored Seminar*, ARI, Alexandria, Virginia, October 1981.

Magen, Z., "Encounter Group Effects on Soccer Team Performance," *Small Group Behavior*, Vol. 7, 1980, pp. 339–344.

Maitland, A. J., *Using Predictor Equations to Assess Potential Performance of Armor Enlistees*, ARI Research Report 1240, Alexandria, Virginia, January 1980.

Maitland, A. J., N. K. Eaton, and J. F. Neff, *Cross-Validation of Predictor Equations for Armor Crewman Performance*, ARI Technical Report 447, Alexandria, Virginia, January 1980.

Maksimova, R. A., "Svyaz' Psikhologicheskogo Klimata v Kollektive s Effektivnostyu Truda" (The Correlation Between Psychological Climate in the Collective and Job Effectiveness), in E. Kuzmin (ed.), *Sotsialnaya i Sotsialnoye Planirovaniye*, University of Leningrad, Leningrad, 1973, pp. 28–33.

Marks, M. L., and P. H. Mirvis, "Environmental Influences on the Performance of a Professional Baseball Team," *Human Organization*, Vol. 40, 1981, pp. 355–360.

Martens, R., and J. A. Peterson, "Group Cohesiveness as a Determinant of Success and Member Satisfaction in Team Performance," *International Review of Sport Sociology*, Vol. 6, 1971, pp. 49–61.

McGrath, J. E., "The Influence of Positive Interpersonal Relations on Adjustment and Effectiveness in Rifle Teams," *Journal of Abnormal and Social Psychology*, Vol. 65, 1962, pp. 365–375.

Medlin, S. M., *Combat Operations Training Effectiveness Analysis Model: 1979 Perspective*, ARI Research Report, Alexandria, Virginia, July 1979.

Meister, D., *Behavioral Foundations of System Development*, John Wiley and Sons, Inc., New York, 1976.

Meliza, L. L., T. D. Scott, and K. I. Epstein, *REALTRAIN Validation for Rifle Squads II: Tactical Performance*, ARI Research Report 1203, Alexandria, Virginia, March 1979.

Melnick, M. J., and M. M. Chemers, "Effects of Group Social Structure on the Success of Basketball Teams," *Research Quarterly*, Vol. 45, 1974, pp. 1–8.

Miller, E. E., and J. M. Bachta, *An Experimental Evaluation of a Tactical Game for Company Level Training*, General Research Corporation, McLean, Virginia, July 1978.

Miller, J. G., *Living Systems*, McGraw-Hill Book Co., New York, 1978.

Mischel, W., *Personality and Assessment*, John Wiley and Sons, Inc., New York, 1968.

Mitchell, T., "Organizational Behavior," *Annual Review of Psychology*, 1979, pp. 243–281.

Moos, R. H., and J. C. Speisman, "Group Compatibility and Productivity," *Journal of Abnormal and Social Psychology*, Vol. 65, 1962, pp. 190–196.

Myers, A., "Team Competition, Success, and the Adjustment of Group Members," *Journal of Abnormal and Social Psychology*, Vol. 65, 1962, pp. 325–332.

Nadler, D. A., "Effects of Feedback on Task Group Behavior: Review of the Experimental Research," *Organizational Behavior and Human Performance*, Vol. 23, 1979, pp. 309–338.

O'Brien, G. E., "The Measurement of Cooperation," *Organizational Behavior and Human Performance*, Vol. 3, 1968, pp. 427–439.

O'Brien, G. E., and F. Harary, "Measurement of the Interactive Effects of Leadership Style and Group Structure Upon Group Performance," *Australian Journal of Psychology*, Vol. 29, No. 1, 1977, pp. 59–71.

O'Brien, G. E., T. T. Hewett, and J. Hornik, "The Effects of Cooperation, Leadership Style, and Member Compatibility Upon Small Group Productivity," *Proceedings of the 20th International Congress of Psychology*, 1972.

O'Brien, G. E., and D. Ilgen, "Effects of Organizational Structure, Leadership Style, and Member Compatibility Upon Small Group Creativity," *Proceedings of the 76th Annual Convention of the American Psychological Association*, Vol. 3, 1968, pp. 555–556.

O'Brien, G. E., and B. Kabanoff, "The Effects of Leadership Style and Group Structure Upon Small Group Productivity: A Test of a Discrepancy Theory of Leader Effectiveness," *Australian Journal of Psychology*, Vol. 33, 1981, pp. 157–168.

O'Brien, G. E., and A. G. Owens, "Effects of Organizational Structure on Correlations between Member Abilities and Group Productivity," *Journal of Applied Psychology*, Vol. 53, 1969, pp. 525–530.

O'Brien, R. E., et al., *Tank Crew (M60A1) Performance Exercise*, ARI Research Product 79-15, Alexandria, Virginia, November 1979.

Obermayer, R. W., and D. Vreuls, *Combat-Ready Crew Performance Measurement System: Phase IIIA Crew Performance Measurement*, Air Force Human Resources Laboratory, Report AFHRL-TR-74-108(IV), Williams AFB, Arizona, December 1974.

Obermayer, R. W., et al., *Combat-Ready Crew Performance Measurement System: Final Report*, Air Force Human Resources Laboratory, Report AFHRL-TR-74-108(1), Williams AFB, Arizona, December 1974.

Porter, L. W., and E. E. Ghiselli, "The Self-Perceptions on Top and Middle Management Personnel," *Personnel Psychology*, Vol. 10, 1957, pp. 397–406.

Pryer, M. W., and B. M. Bass, "Some Effects of Feedback on Behavior in Groups," *Sociometry*, Vol. 22, 1959, pp. 56–63.

Rapoport, A., "From N=2 to N≥3," *Contemporary Psychology*, Vol. 16, 1971, pp. 49–51.

Rapoport, A., J. P. Kahan, and W. E. Stein, "Decisions of Timing in Experimental Probabilistic Duels, *Journal of Mathematical Psychology*, Vol. 13, 1976, pp. 163–191.

Reddy, W. B., "Diagnosing Team Problem-Solving Effectiveness: A Comparison of Four Populations," *Small Group Behavior*, Vol. 6, 1975, pp. 174–186.

Reddy, W. B., and A. Byrnes, "Effects of Interpersonal Group Composition on the Problem-Solving Behavior of Middle Managers," *Journal of Applied Psychology*, Vol. 56, 1972, pp. 516–517.

Rizzo, W. A., "Navy Team Training: Some Critical Issues," in S. E. Goldin and P. W. Thorndyke (eds.), *Improving Team Performance: Proceedings of The Rand Team Performance Workshop*, The Rand Corporation, R-2606-ONR, August 1980.

Roberts, J. M., Q. S. Meeker, and J. C. Aller, "Action Styles and Management Game Performance: An Exploratory Consideration," *Naval War College Review*, Vol. 24, 1972, pp. 65–81.

Roberts, K. H., "Some Thoughts About Effective Team-Effectiveness Research," in S. E. Goldin and P. W. Thorndyke (eds.), *Improving Team Performance: Proceedings of The Rand Team Performance Workshop*, The Rand Corporation, R-2606-ONR, August 1980.

Rohde, K. J., "Theoretical and Experimental Analysis of Leadership Ability," *Psychological Reports*, Vol. 4, 1958, pp. 243–278.

Root, R. T., et al., *Tactical Engagement Simulation Training: A Method for Learning the Realities of Combat*, ARI Technical Paper 370, Alexandria, Virginia, August 1979.

Rosenberg, S., and R. L. Hall, "The Effects of Different Social Feedback Conditions Upon Performance in Dyadic Teams," *Journal of Abnormal and Social Psychology*, Vol. 57, 1958, pp. 271–277.

Ruscoe, G. C., "Application of Living Systems Process Analysis to Army Organizations: A Review and Critique," *Behavioral Science*, Vol. 27, 1982, pp. 203–213.

Ryan, T. G., and L. G. Yates, *Report of Exercise Observations: Operational Readiness Training Tests (ORTT)*, ARI Research Memorandum 77-7, Alexandria, Virginia, December 1977.

Schutz, W. C., *FIRO: A Three-Dimensional Theory of Interpersonal Behavior*, Holt, Rinehart, & Winston, New York, 1958.

Scott, T. D., L. L. Meliza, G. D. Hardy, Jr., and J. H. Banks, *Armor/Anti-Armor Team Tactical Performance*, ARI Research Report 1218, Alexandria, Virginia, July 1979.

Scott, T. D., L. L. Meliza, G. D. Hardy, Jr., J. H. Banks, and L. E. Word, *REALTRAIN Validation for Armor/Anti-Armor Teams*, ARI Research Report 1204, Alexandria, Virginia, March 1979.

Shalinsky, W., "Group Composition as a Factor in Assembly Effects," *Human Relations*, Vol. 22, No. 5, October 1969, pp. 457–464.

Shavelson, R. J., and N. M. Webb, "Generalizability Theory: 1973–1980," *British Journal of Mathematical and Statistical Psychology*, Vol. 34, 1981, pp. 133–166. (Also Rand P-6580.)

Shaw, M. E., "A Note Concerning Homogeneity of Membership and Group Problem Solving," *Journal of Abnormal and Social Psychology*, Vol. 60, 1960, pp. 448–450.

Shaw, M. E., *Group Dynamics: The Psychology of Small Group Behavior*, McGraw-Hill Book Co., New York, 1971.

Shaw, M. E., and J. M. Blume, "Effects of Leadership Style Upon Group Performance as a Function of Task Structure," *Journal of Personality and Social Psychology*, No. 66, 1966, pp. 238–242.

Siegel, A. I., and P. J. Federman, "Communications Content Training as an Ingredient in Effective Team Performance," *Ergonomics*, Vol. 16, 1973, pp. 403–416.

Simonton, D. K., "Land Battles, Generals, and Armies: Individual and Situational Determinants of Victory and Casualties," *Journal of Personality and Social Psychology*, Vol. 38, No. 1, January 1980, pp. 110–119.

Skinner, B. F., "The Steep and Thorny Way to a Science of Behavior," *American Psychologist*, Vol. 30, 1975, pp. 42–49.

Smith, K. H., "Changes in Group Structure through Individual and Group Feedback," *Journal of Personality and Social Psychology*, Vol. 24, No. 3, December 1972, pp. 425–428.

Sorenson, J. R., "Group Member Traits, Group Process, and Group Performance," *Human Relations*, Vol. 26, 1973, pp. 639–655.

Spoelders-Claes, R., "The Effect of Varying Feedback on the Effectiveness of a Small Group on a Physical Task," *Psychologica Belgica*, Vol. 13, 1973, pp. 61–68.

Steiner, I. D., *Group Process and Productivity*, Academic Press, New York, 1972.

Stogdill, R. M., *Handbook of Leadership: A Survey of Theory and Research*, Collier-Macmillan, London, 1974.

Stone, T. H., "Effects of Mode of Organization and Feedback Level on Creative Task Groups," *Journal of Applied Psychology*, Vol. 55, No. 4, August 1971, pp. 324–330.

Straub, W. F. (ed.), *Sport Psychology: An Analysis of Athlete Behavior*, Mouvement Publications, Ithaca, New York, 1978.

Straub, W. F., "Team Cohesion in Athletics," *International Journal of Sport Psychology*, Vol. 6, 1975, pp. 125–133.

Sulzen, R. H., *The Effects of Repeated Engagement Simulation Exercises on Individual and Collective Performance*, ARI Technical Report 485, Alexandria, Virginia, August 1980.

Sventsitskiy, A. L., "Sotsialno-Psikhologicheskiy Klimat Pervichnogo Proizvodstvennogo Kollektiva Kak Obyekt Issledovaniya" (The Social Psychological Climate of the Primary Work Collective as an Object of Investigation), in E. Kuzmin (ed.), *Sotsialnaya Psikhologiya i Sotsialnoye Planirovaniye*, University of Leningrad, Leningrad, 1973, pp. 20–28.

Terborg, J. R., C. Castore, and J. A. DeNinno, "Longitudinal-Field Investigation of Impact of Group Composition on Group Performance and Cohesion," *Journal of Personality and Social Psychology*, Vol. 34, No. 5, November 1976, pp. 782–790.

Thorndyke, P. W., and M. G. Weiner, *Improving Training and Performance of Navy Teams: A Design for a Research Program*, The Rand Corporation, R-2607-ONR, July 1980.

Triandis, H. C., E. R. Hall, and R. B. Ewen, "Member Heterogeneity and Dyadic Creativity," *Human Relations*, Vol. 18, No. 1, February 1965, pp. 33–55.

Triandis, H. C., "Categories of Thought of Managers, Clerks, and Workmen about Jobs and People in an Industry," *Journal of Applied Psychology*, Vol. 43, No. 5, October 1959b, pp. 338–344.

Triandis, H. C., "Cognitive Similarity and Communication in a Dyad," *Human Relations*, Vol. 13, No. 2, May 1960a, pp. 175–183.

Triandis, H. C., "Cognitive Similarity and Interpersonal Communication in Industry," *Journal of Applied Psychology*, Vol. 43, No. 5, October 1959c, pp. 321–326.

Triandis, H. C., "Differential Perception of Certain Jobs and People by Managers, Clerks, and Workers in Industry," *Journal of Applied Psychology*, Vol. 43, No. 4, August 1959a, pp. 221–225.

Triandis, H. C., "Some Determinants of Interpersonal Communication," *Human Relations*, Vol. 13, No. 3, August 1960b, pp. 279–287.

Tuckman, B. W., "Group Composition and Group Performance of Structured and Unstructured Tasks," *Journal of Experimental Social Psychology*, Vol. 3, 1967, pp. 25–40.

Tuckman, B. W., "Personality Structure, Group Composition, and Group Functioning," *Sociometry*, Vol. 27, 1964, pp. 469–487.

Turney, J. R., and S. L. Cohen, *Defining the Nature of Team Skills in Navy Team Training and Performance*, General Physics Corporation, Report GP-R-43017-2, Columbia, Maryland, September 1981.

Turney, J. R., S. L. Cohen, and L. Greenberg, *Targets for Team Skills Training*, General Physics Corporation, Report GP-R-43017, Columbia, Maryland, April 1981.

Underwood, W. J., and L. J. Krafft, "Interpersonal Compatibility and Managerial Work Effectiveness," *Journal of Applied Psychology*, Vol. 58, No. 1, August 1973, pp. 89–94.

Van Zelst, R. H., "Validation of Sociometric Regrouping Procedure," *Journal of Abnormal and Social Psychology*, Vol. 47, 1952, pp. 299–301.

Vos, K., and W. Brinkman, "Succes en Cohesie in Sportgroepen" (Success and Cohesion in Sports Teams), *Sociologische Gios*, Vol. 14, 1967, pp. 30–40.

Walter, G. A., and R. E. Miles, "Essential Elements for Improving Task-Group Membership Behaviors," *Proceedings of the 80th Annual Convention of the American Psychological Association*, Vol. 7, 1972, pp. 461–462.

Weber, R. J., "Effects of Process Feedback, Consultation, and Knowledge of Results on Perceptions of Group Process," *Proceedings of the 80th Annual Convention of the American Psychological Association*, Vol. 7, 1972, pp. 459–460.

Weber, R. J., "Effects of Videotape Feedback on Task Group Behavior," *Proceedings of the 79th Annual Convention of the American Psychological Association*, Vol. 6, 1971, pp. 499–500.

Widmeyer, W. N., "When Cohesiveness Predicts Performance Outcome in Sport," *Dissertation Abstracts International*, Vol. 38 (6-B), 1977, pp. 2951–2952.

Widmeyer, W. N., and R. Martens, "When Cohesion Predicts Performance Outcome in Sport," *Research Quarterly*, Vol. 49, 1978, pp. 372–380.

Widmeyer, W. N., J. W. Loy, and J. M. Roberts, "The Relative Contributions of Action Styles and Ability to the Performance Outcomes of Doubles Tennis Teams," *Psychology of Motor Behavior and Sport*, 1979, pp. 209–218.

Wiest, W. M., L. W. Porter, and E. E. Ghiselli, "Relationship between Individual Proficiency and Team Performance and Efficiency," *Journal of Applied Psychology*, Vol. 45, No. 6, December 1961, pp. 435–440.

Yukl, G. A., and D. D. van Fleet, "Cross-situational, Multimethod Research on Military Leader Effectiveness," *Organizational Behavior and Human Performance*, Vol. 30, 1982, pp. 87–108.

Zajonc, R. B., "The Effects of Feedback and Probability of Group Success on Individual and Group Performance," *Human Relations*, Vol. 15, 1962, pp. 149–161.

Zander, A., "Students' Criteria of Satisfaction in a Classroom Committee Project," *Human Relations*, Vol. 22, 1969, pp. 195–207.

Zander, A., *Motives and Goals in Groups*, Academic Press, New York, 1971.

Zander, A., "Motivation and Performance of Sports Groups," in W. F. Straub (ed.), *Sport Psychology: An Analysis of Athlete Behavior*, Mouvement Publications, Ithaca, New York, 1978.

Zander, A., and H. Medow, "Strength of Group and Desire for Attainable Group Aspirations," *Journal of Personality*, Vol. 33, 1965, pp. 122–139.

Zander, A., and D. Wolfe, "Administrative Rewards and Coordination among Committee Members," *Administrative Science Quarterly*, Vol. 9, 1964, pp. 50–69.